CONTENTS

SERIES A   VOLUME 181

No. A 984.   24 September 1942

No. A 985.   31 December 1942

Contents

The hardness of primary solid solutions with special reference to alloys of silver

By John H. Frye (Jr) and William Hume-Rothery, F.R.S.

(Received 4 December 1941)

Accurate Meyer analyses have been made on annealed bars of vacuum-melted spectro-scopically pure silver, assay silver, and on binary alloys of silver containing up to 5 atomic % of cadmium, indium, tin, antimony, zinc, aluminium, magnesium, and gold. The results show that in contradiction to the statements of some investigators, the ultimate Meyer hardness P_u increases as the grain size of the specimens diminishes. By means of suitable annealing treatments results were obtained for different grain sizes of alloys containing approximately 2·4 and 5·0 atomic % of the above elements, and these were used to deduce P_u values referring to one standard grain size. These values enable an accurate comparison of the relative hardening effects of the different solutes to be made. The results are discussed from the standpoint of lattice distortion, and it is shown that for a given atomic percentage of solute, the increase in P_u is proportional to the square of the lattice distortion for solutes in the same row of the Periodic Table. No simple relation exists for solutes from different rows of the Periodic Table, and for a given solute the relation between P_u and the composition is not necessarily a simple one. The conclusions are tested by experiments on ternary alloys, and data for copper alloys are discussed.

1. Introduction

It is now well known that the majority of pure metals are comparatively soft, but that in many cases the hardness can be increased by alloying with one or more elements which enter into primary substitutional solid solution. Although the hardness of these solid solutions is seldom very great, its study is of considerable interest, for they constitute the simplest type of alloy, and in many cases the alloys of complex structure contain a primary solid solution as the general matrix. In developing a theory of the hardness of alloys, the study of primary·solid solutions is thus of great importance, and a large amount of published data already exists. Unfortunately, most of this is unsuitable for critical comparison, for many different hardness tests are in use, and, although for certain alloys empirical relations have been found to connect the results of different methods of testing, it is seldom that the results of one investigator can be compared accurately with those of another. To some extent this difficulty lies in the nature of the properties summarized by the term 'hardness'. If, for example, hardness is defined as resistance to permanent deformation, or to permanent indentation, the nature of the mechanical process has to be considered, and it does not follow that the same physical property is measured when indentations are made with differently shaped indenting tools. In the popular Brinell test, it is well known that the hardness number of a specimen usually varies with the load and the ball diameter, and comparison of isolated Brinell numbers may be quite misleading. Apart from the nature of the test employed, the hardness of an alloy may be influenced by factors such as grain size,

and unless these are controlled, comparisons may again be misleading. To some extent any hardness test is arbitrary, but we shall here accept the view of O'Neill (1934) that the method of Meyer (1908) is the most satisfactory way of expressing results obtained with spherical indenters, and that the ultimate Meyer hardness, P_u, is the least unsatisfactory way of indicating hardness by means of a single constant. This method depends on the fact that for many substances the relation between the load L, and the diameter of impression d, for a ball of given diameter is expressed by the Meyer equation

$$L = ad^n.$$

Here n and a are constants for a given material, and d is the diameter of the indentation after recovery. The so-called Meyer hardness, P, at a particular load L, is given by the expression

$$P = 4L/\pi d^2,$$

where $\frac{1}{4}\pi d^2$ represents the projected area of the impression. Combination of these two expressions leads to the relation

$$P = 4ad^{n-2}/\pi.$$

The usual method of making a Meyer analysis is to plot $\log L$ against $\log d$, when a straight line is obtained over the range $0 \cdot 1 < d/D < 1$, where D is the diameter of the ball. The slope of this line gives the constant n, and if the straight line is produced to the point where $d = D$, the corresponding load may be used to calculate the so-called ultimate Meyer hardness, P_u, given by

$$P_u = (4a/\pi)\, D^{n-2}.$$

The constant a is the value of L when $d = 1$; the corresponding Meyer hardness, P_a, is given by

$$P_a = 4a/\pi.$$

The Meyer hardness P_a has no fundamental significance, but with reasonably large balls it is a convenient measure of hardness at small degrees of deformation *provided that a ball of constant diameter is always used*; values of P_a obtained with balls of different diameter cannot be compared. The ultimate Meyer hardness, P_u, is of much greater significance, and is only influenced slightly by the ball diameter. Some investigators have also concluded that P_u is independent of the grain size of the material, but this has not been confirmed by the present work with a 4 mm. ball, and we have therefore been compelled to take additional precautions in order to compare the results for specimens with a constant grain size.

The only detailed and really satisfactory work on the relative hardness of different solid solutions appears to be that of Norbury (1923), who carried out Meyer analyses for the solid solutions of a number of elements in copper, and showed that with a few exceptions there was a clear correspondence between the increase in hardness, and the increase in atomic volume produced by 1 atomic % of the different solutes. In the present work we have first investigated the relation

between grain size and the ultimate Meyer hardness of copper, silver, and of a number of silver alloys, and have shown that the P_u values are slightly influenced by grain size. We have then carried out complete Meyer analyses on solid solutions of silver containing approximately 2·4 and 5·0 atomic % of cadmium, indium, tin, antimony, zinc, magnesium, aluminium, and gold,* and also for ternary solid solutions with zinc and cadmium. In view of the scarcity of accurate hardness data, we have investigated the above alloys in great detail, rather than attempted to cover a wider range in a less accurate way. We have obtained results referring to one standard grain size, and in this way we hope that a really reliable series of P_u values has been obtained, and we have then discussed the results in the light of the lattice distortion to be expected in this type of alloy. The experimental methods and results are described in §§ 2 and 3 respectively, and the discussion is contained in § 4.

2. Experimental methods

The alloys used were supplied by Messrs Johnson Matthey and Co., Ltd., and were made from high-purity metals. They were cast into graphite moulds 1½ in. in diameter, and after surfacing in a lathe, were rolled to 1¼ in. diameter, after which they were annealed for 4 days at 650° C in iron tubes packed with charcoal, except for the alloys with magnesium, which were annealed in evacuated tubes. The rods were then rolled to ¾ in. diameter, and were received by us in this form.

The silver used for the preparation of the alloys was described as being of 99·99 + % purity, the principal impurity being oxygen. In view of the results described later, it was considered desirable also to measure the hardness of spectroscopically pure silver, melted and cast in vacuo, and this material was cast in the form of a bar 1¾ in. in diameter which was machined to 1½ in., cold-rolled to 1¼ in., annealed for 8 hr. in a vacuum at 650° C, and finally cold-rolled to ¾ in. diameter rod. The remaining metals were of high purity, and the analyses of the rods indicated that no contamination occurred during preparation. No signs of mechanical unsoundness (cracks, blowholes, etc.) were noted, and the good reproducibility of the results indicated that the specimens were entirely satisfactory. Our thanks are due to Mr A. R. Powell for the great care he took over this work in the laboratories of Messrs Johnson Matthey and Co., Ltd.

Preparation of specimens

Specimens of about 2 cm. long were cut from the bars by parting in a lathe, or by sawing, and were rubbed down on steel files, and polished on no. 1 Hubert emery paper. The specimens were tested with a spirit level, and were reground if the angle between the faces exceeded 2°. They were then annealed in sealed evacuated glass tubes, and only specimens from the same rod were annealed in the same tube. This precaution is very necessary, because even with what would

* Only one alloy with gold was examined.

usually be classed as non-volatile metals, remarkable changes in composition may occur if different alloys are annealed in the same tube. The time and temperature of the final annealing treatment were adjusted so as to produce the desired grain size, and were always sufficient to ensure complete recrystallization.

Chemical analysis of specimens

After the hardness tests had been carried out, the composition of the specimens was determined by taking drillings from the immediate vicinity of the hardness impression. For each alloy system the 5 atomic % alloy* was also tested by the analysis of a sample taken from the periphery in order to detect radial segregation. This type of segregation was fortunately very small, and was usually of the order 0·02–0·03 % by weight, and as there seems to be no obvious method of correcting the hardness results for an effect of this kind, this source of error was neglected and the composition was taken to be given by the analysis of the material surrounding the impression. Longitudinal segregation was guarded against by the analysis of two samples taken from widely separated positions in the bar. If the two analyses were in agreement, it was assumed that longitudinal segregation was absent, but, if any appreciable difference was found, additional analyses were carried out, and this was also done as a further check in some cases even where the two analyses agreed. Longitudinal segregation was found to be most serious in the 5 % silver-tin alloy for which the tin content was found to vary from 5·25 to 5·66 % by weight at different places.

Most of the analytical work was carried out by Messrs Johnson Matthey and Co., Ltd., and the remainder by the Southern Testing Laboratories of Birmingham, Ala., U.S.A.

Hardness testing methods

Examination of previous data suggested that satisfactory results would be obtained on ¾ in. diameter bars† by the use of a ball of diameter 4 mm. under loads up to 500 kg. The hardness tests in England were carried out on two machines, the first being a standard Brinell machine kindly placed at our disposal by Professor R. V. Southwell, F.R.S., at the Oxford University Engineering Laboratory. This was provided with a specially light yoke, and by the addition of suitable weights, loads of approximately 100, 200, 300, 400, and 500 kg. could be obtained. The actual loads were obtained by calibration against standard steels kindly provided by the National Physical Laboratory. Loads of approximately 20 and 40 kg. were obtained by means of a simple lever machine made by the Metropolitan Vickers Electrical Company, Ltd., and the exact loads were again determined by calibration against National Physical Laboratory standard steels. Our thanks are due to Dr H. J. Gough, F.R.S., for his help in connexion with the calibrations. For all the work except that on spectroscopically pure silver, the Meyer analyses

---

* The segregation would naturally be very much less for the 2·4 atomic % alloys.

† The use of larger bars and a 10 mm. ball was considered, but was impracticable on account of the expense involved.

were made from observations at six different loads, the exact values being 19·5, 39·2, 97·3, 195, 297, and 391 kg.

Owing to the outbreak of war, and the return of one of us (J. H. F.) to America, the work on the spectroscopically pure silver was carried out at Birmingham, Ala., U.S.A. A 5000 lb. capacity tensile and compressive tester was kindly made available by the American Cast Iron Pipe Company, and was satisfactorily adapted for the hardness tests. Elaborate precautions* were taken to ensure that the results were comparable with those obtained at Oxford, and the loads used were calibrated against a standard weight certified by the U.S. Bureau of Standards. The Meyer analyses were made from observations at ten different loads between 100 and 1000 lb.

In all the work the Meyer analyses were made by the method of concentric impressions, as in the work of Meyer and of Norbury. Experiments with alloys containing 5 atomic % of zinc and aluminium showed no difference in the diameters of the impressions obtained after application of loads for times varying from 1 to 5 min., and the duration of the load in the present experiments was never less than 1 min. After each application of the load, the impression was measured to an estimated 0·001 mm. for at least two mutually perpendicular diameters, and in many cases other diameters at 45° were also measured. The measurements were made with a travelling microscope kindly lent by Professor F. Soddy, F.R.S., and the accuracy was tested by comparing measurements of the same impression made at Oxford and at the National Physical Laboratory. For the work in America the travelling microscope used was calibrated against a line standard from the National Bureau of Standards.

Grain-size measurement and standardization

After the measurement of the impressions, one specimen of each alloy from each heat treatment was sectioned, and the section was polished and etched, and used for the determination of grain size, which was carried out by means of a Vickers Projection Microscope. From 150 to 200 individual grains were always counted, and in cases where there was any considerable variation in the grain size as many as 600 grains were counted on a single specimen. Twin crystals were not counted as separate crystals. This decision is admittedly arbitrary, but as no evidence was found of any marked variation in the number of twins per crystal, it does not affect the main purpose of the grain-size measurement, which was to obtain values for a standard grain size.

The grain size was expressed in grains per square millimetre, and a standard grain size of 59 grains/mm.² was chosen for the comparison of results. This grain size is considerably smaller than that required in order to obtain uniform impressions, and is such that both larger and smaller grain sizes can be obtained by suit-

---

* These details will be found in the thesis 'The hardness of certain primary metallic solid solutions', presented by J. H. Frye for the degree of Doctor of Philosophy in the University of Oxford.

able annealing treatment. For each binary alloy, the hardness of at least two grain sizes was examined, one of which was made as near as possible to the standard value of 59 grains/mm.².* According to Angus & Summers (1925), the hardness of copper, and of α-bronze, is a linear function of the grain-boundary area per unit volume, which was defined as $3\sqrt{q}$, where q is the number of grains per square millimetre, and thus involves the assumption of cubical grains. We have used this relation to correct our hardness figures to the standard grain size of 59 grains/mm.². In some cases the problem was complicated by the fact that the specimens of the same rod with different grain sizes had slightly different compositions, and here a method of successive approximations was used to allow for the simultaneous variation of grain size and composition in order to refer to the standard value.

3. EXPERIMENTAL RESULTS

A. *The effect of grain size on the ultimate Meyer hardness*

Previous investigators had suggested that the ultimate Meyer hardness P_u was independent of the grain size, but the present work did not support this conclusion, and the question was therefore examined in detail for copper, silver, and also for the silver alloys. The results for the pure metals are given in table 1, from which it will be seen that a slight but clear effect exists, and that the values of P_u increase as the grain size diminishes.

TABLE 1

metal	grains/mm.²	ultimate Meyer hardness P_u. 4 mm. ball applied for not less than 60 sec.	Meyer constant n
copper	0·25	61·4	2·543
	560	64·7	2·511
	1070	66·7	2·533
silver, 99·99 + %	0·12	43·4	2·644
	2420	48·0	2·455
silver, spectroscopically pure vacuum method	1·5	42·4	2·709
	155	43·1	2·548
	407	44·4	2·528

In the case of the alloys, the position was complicated by the fact that specimens from the same bar but with different grain sizes sometimes had slightly different compositions. Table 2 summarizes the data for the cases in which the analyses of the two specimens agreed very closely, and it will be seen that in every case the value of P_u increases as the grain size becomes smaller. The error in P_u is regarded

* In general one of the two grain sizes was very near to the standard value, the only exception being the silver-gold alloy, for which the grain sizes of the two specimens were 13 and 550 grains/mm.² respectively. The corresponding values of P_u were 46·3 and 50·1, from which it will be appreciated that the correction was nearly always very small.

as less than 0·5, and from the figures given in table 2 it will be appreciated that with smaller variations in grain size there were many cases in which the effect of grain size was within the experimental error, and we are not therefore submitting these results for detailed discussion. The data may be summarized by saying that in the whole of the forty-four alloys which were submitted to both hardness tests

TABLE 2

alloy system	composition of alloy. Weight % Ag	grains/mm.²	ultimate Meyer hardness P_u. 4 mm. ball applied for not less than 60 sec.	Meyer constant n
Au–Ag	91·21	13	46·3	2·732
	91·24	550	50·1	2·674
Sn–Ag	94·75	9	55·7	2·679
	94·77	48	59·7	2·694
Zn–Ag	96·82	3·9	46·9	2·641
	96·92	84	48·4	2·631
Zn–Ag	98·45	49	46·6	2·664
	98·51	123	47·1	2·651
In–Ag	97·32	44	46·9	2·683
	assumed constant	2180	53·8	2·505
Al–Ag	99·32	65	47·4	2·666
	99·25	249	49·9	2·619
Al–Ag	98·95	43	51·2	2·697
	98·97	215	52·4	2·601
Mg–Ag	98·75	4·9	48·0	2·578
	98·77	92	48·5	2·550
Mg–Ag	0·58 % Mg	44	44·2	2·616
	0·61 % Mg	1720	48·0	2·591

and chemical analysis, only four pairs were found in which a decrease in grain size was accompanied by a decrease in P_u, and in all of these the difference was within the experimental error. The evidence appears therefore to be conclusive that P_u varies with the grain size, and that for accurate comparison the hardness numbers must refer to a standard grain size.

As will be seen from table 1, the Meyer constant n decreases with decreasing grain size for both the high-purity silver and for the spectroscopically pure metal. This is in agreement with previous results for other metals, and as will be seen from table 2 the same characteristic is shown by the alloys. The differences are usually small, and of the same order as the experimental error, but the general tendency is unmistakable.* In the case of copper the largest grain size has the largest value of n in agreement with the above principle, but the remaining points are in the reverse order, although the differences are all very small.

* In table 2 the only exception is shown by the silver-tin alloys and the difference in n is only 0·015.

B. *The relative hardening effects of different solutes*

In order to present the data as concisely as possible we are submitting only one table containing the final results, all referring to a standard grain size of 59 grains/mm.^{2}. For many alloys two or more specimens were examined from each heat treatment. In cases where the analysis indicated that the bars were of uniform composition, the average values were taken and were used to obtain the value referring to the standard grain size. In cases where the compositions of the different specimens varied slightly, we have only included one value in table 3, and this represents an average of both the composition and hardness determinations. It must be emphasized that although table 3 shows only a limited number of results, these are average values based on the examination of more than 100 specimens, each involving a complete Meyer analysis. The values of P_u are accurate to within ± 0.5 kg./mm.^{2}. The accuracy of the P_a values is less easy to estimate, since, although the experimental error is greater for the smaller impressions, this source of error is to some extent compensated by the fact that P_a is determined from the best straight line through the whole series of points in the Meyer analysis, and we consider that the P_a values can certainly be relied upon to within ± 1.0 kg./mm.

TABLE 3. RESULTS FOR SILVER ALLOYS

atomic % solute		$\Delta a \times 10^3$ (A)	$(\Delta a)^2 \times 10^7$	P_a kg./mm.^{2}	P_u kg./mm.^{2}	n
2·64	Cd	+ 4·9	2·40	19·9	44·3	2·58
4·71	Cd	+ 9·2	8·46	23·0	47·3	2·52
7·59	Cd	+15·5	—	21·1	47·6	2·59
2·54	In	+ 7·2	5·18	19·1	46·4	2·64
5·01	In	+15·0	22·5	21·7	52·7	2·64
2·36	Sn	+ 9·4	8·84	21·0	48·1	2·60
4·98	Sn	+20·4	41·6	22·7	60·7	2·71
2·11	Sb	+12·3	15·1	22·7	52·5	2·60
2·52	Mg	+ 2·0	0·40	18·9	44·3	2·61
5·24	Mg	+ 4·5	2·03	22·3	48·3	2·56
2·59	Al	− 3·25	1·056	18·4	47·0	2·67
4·15	Al	− 5·2	2·70	21·0	51·6	2·65
5·15	Al	− 6·5	4·23	20·5	52·6	2·68
2·53	Zn	− 4·3	1·85	18·6	46·7	2·66
4·98	Zn	− 8·5	7·23	20·0	48·0	2·63
5·00	Au	− 1·5	0·225	17·4	47·1	2·72
Ag, 99·99 +		—	—	19·0	44·1	2·61
Ag, spectroscopically pure vacuum melted				18·4	42·8	2·61

4. DISCUSSION

In attempting to correlate the hardening effects of different elements in solid solutions, we have to bear in mind that in an indentation test the deformation is highly localized, so that the metal in the neighbourhood of the impression has

undergone extremely intense deformation. Recent work has shown that in the early stages of the deformation of a metal, the process involves gliding and breaking up into crystallites, whilst with extreme deformation, an actual distortion of the lattice is indicated by the X-ray results. We may therefore expect lattice distortion to be intimately connected with the hardening process, since, in the early stages of the deformation, the distortion will affect the ease of the gliding process, whilst, in the later stages of the deformation, the lattice distortion produced by the deformation process will be affected by that already existing owing to the presence of the solute atoms. For copper alloys, this general kind of relation was shown clearly by the work of Norbury, some of whose results are shown in figure 1, in

difference in size of solvent and solute atom
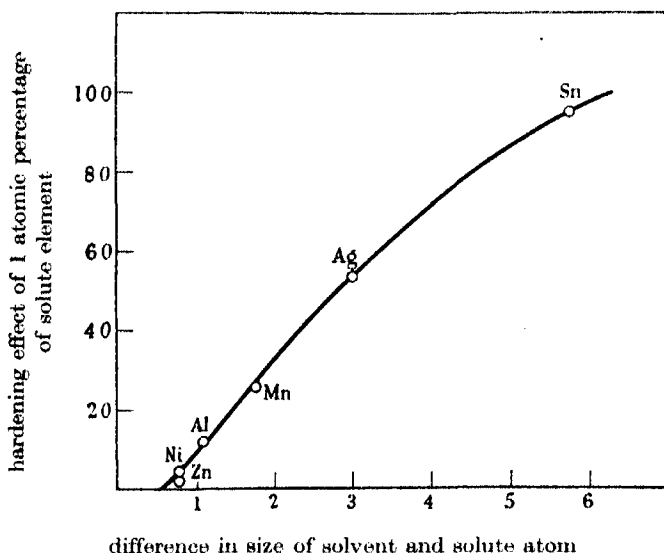
FIGURE 1

which the hardening effects of the different solvent elements are plotted against the difference in size of solvent and solute atoms, this difference being estimated from measurements of the densities of the solid solutions. The result for the silicon alloy was clearly anomalous, and this was regarded as due to the formation of compounds. More recent work has shown that the solubility of silicon in copper is small at room temperature, and it is possible that the anomalous results are due partly to the breakdown of the supersaturated solid solution during the test; for the same reason the point for silver is not entirely free from doubt. These results do, however, show clearly the general effect of atomic volumes on the relative hardening produced by different solutes.

To carry the matter further it is necessary to consider the nature of the lattice distortion in a solid solution. It is well known that the lattice spacing of a solid solution as indicated by X-rays represents a mean value only, and that regions of

more intense distortion exist in the neighbourhood of each solute atom. It is also highly probable that even in dilute solid solutions there is always a certain degree of local order, in the sense that solute atoms tend to keep far away from one another. A dilute primary solid solution may thus be regarded as containing a number of centres of distortion, and it would seem probable, therefore, that in comparing the relative hardening effects of different solutes, we should at first compare only alloys with approximately the same percentage of solute atoms. There is no reason why the hardening effect should be a simple function of the composition, since the mechanism of hardening may well be different for different percentages of the solute; this applies particularly to alloys with short-range order, since, in these, certain critical compositions may correspond with a sudden increase in the number of places where two solute atoms are first, second, third, ..., etc., closest neighbours. If the data in table 3 are examined, it will be seen that, when compared with the values for spectroscopically pure silver,* the ultimate Meyer hardness is increased by the addition of 2·6 and 4·7 atomic % of cadmium, whereas the ultimate Meyer hardness of the alloy containing 7·6 atomic % of cadmium is little greater than that of the 4·7 atomic % alloy, whilst the P_u value is actually less. The mechanism of hardening appears therefore to change between 5 and 7·6 atomic % cadmium, and for the remainder of this paper we shall confine out attention to the alloys containing roughly 2·4 and 5 atomic % of solute.

In figure 2 the P_u values for the alloys with cadmium, indium, tin, antimony, and magnesium are plotted against the lattice distortion (Δa) in the alloys concerned. The lattice distortions at the exact compositions were estimated from the data of Hume-Rothery, Lewin & Reynolds (1936), and of Owen & Roberts (1939), and are included in table 3, which also shows the values of $(\Delta a)^2$. The values for the magnesium alloys are from new data which are to be published shortly.

The hardening effect of magnesium is clearly greater than that of the remaining elements, and this may be connected with the higher electrochemical factor in the magnesium-silver system, which produces a stronger binding between the solvent and solute atoms. For the remaining elements the points lie roughly on a straight line which does not, however, pass through the value for the vacuum-melted silver. The agreement with a linear relation is not, however, very good, and in figure 3 a and b the same values of P_u are plotted against the squares of the lattice distortions, and more interesting relations are shown. The points for the 2·4 atomic % alloys with cadmium, indium, tin, and antimony lie accurately on one straight line which passes through the point
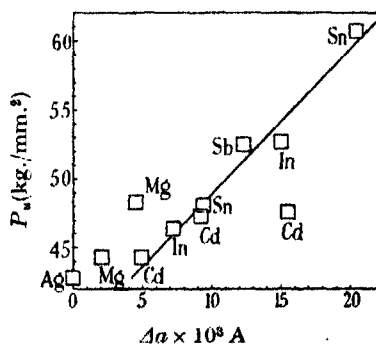
FIGURE 2

* Cadmium acts as a deoxidizer, so that the comparison should be made with the spectroscopically pure vacuum-melted metal.

for the vacuum-melted silver. The points for the 5 atomic % alloys with cadmium, indium, and tin lie on a second straight line also passing through the point for vacuum-melted silver, but of lower gradient. We may therefore conclude that for this series of alloys, at a given small atomic percentage of solute, the increase in P_u is proportional to the square of the lattice distortion.
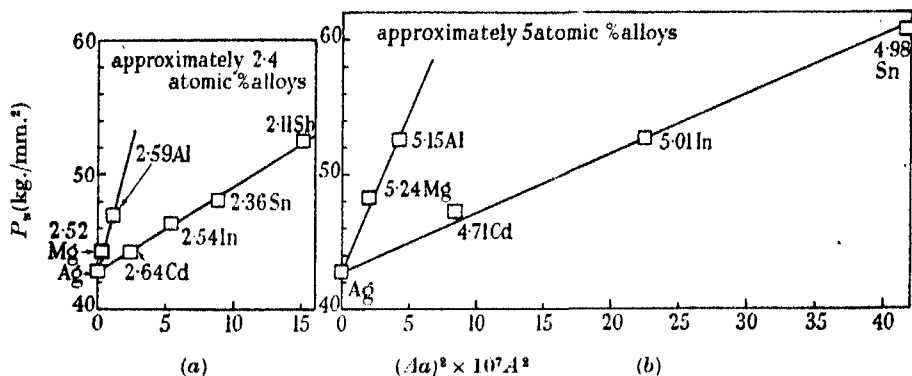
FIGURE 3

The points in figure 2 all refer to alloys in which the formation of the solid solution is accompanied by an expansion. Aluminium is in the same horizontal row of the Periodic Table as magnesium but produces a slight lattice contraction of silver. It is interesting to note that, in spite of the opposite sign of the lattice distortion, the increase in hardness at a given atomic percentage of magnesium and aluminium is again directly proportional to the square of the lattice distortion; this is shown in figure 3 *a* and *b*. Magnesium and aluminium are both very electropositive, and the electrochemical factors, although different for the two systems silver-magnesium and silver-aluminium, are more nearly similar than in the systems silver-magnesium and silver-cadmium.

These results suggest clearly that for a given small atomic percentage of solutes which are adjacent members of the same row of the Periodic Table, the increase in hardness of silver is proportional to the square of the lattice distortion. No simple relation appears to exist between the relative effects of solutes in different rows of the Periodic Table, and the P_u values for the silver-zinc alloys are not simply related to those for the magnesium, aluminium, or the cadmium, indium, ... series. The greatest increase in hardness per unit lattice distortion is shown in the system silver-gold, and it is perhaps significant that the *ion* of gold is larger than that of silver, whereas the ions of all the remaining solutes are smaller. These results suggest therefore that the mean lattice distortions by themselves are not simply related to the relative hardening effects of the different solutes. With a series of elements such as cadmium, indium, ... the underlying ions of the solute are of the same structure, the quantum numbers of the valency electrons are identical, and the details of the lattice distortion are presumably sufficiently alike for the mean

distortion to be simply related to the hardening effect. With another series of elements as solutes, the mechanism of the distortion may be different, so that the relation between hardness and lattice distortion is no longer the same.

In order to test this point, a few experiments were made on ternary alloys of silver with cadmium and zinc. Since the lattice of silver is expanded by cadmium and contracted by zinc, ternary alloys may be made in which the mean lattice distortion is very small. If the hardness of the solid solution depended only on the mean lattice distortion, and not on the details of the distortion process, these alloys should be only very slightly harder than pure silver, whereas as can be seen from table 4 the hardening effect is considerable. Examination of the data in table 4 shows in fact that the P_u values of the two ternary alloys are an additive function, i.e. the hardness P_u of the dilute ternary solid solution containing x atomic % cadmium and y atomic % zinc is given by adding to the hardness of pure silver the increases in hardness found in binary alloys containing x and y atomic % of cadmium and zinc respectively. This appears of interest, since it implies that although, for the mean lattice distortion, the expansion produced by cadmium may be cancelled by the contraction produced by zinc, the localized

TABLE 4

composition of ternary alloy	Meyer constant n	P_u	P_u calculated by assuming increase in P_u is additive
1·15 at % Cd 1·58 at % Zn	2·624	45·7	45·9
3·01 at % Cd 2·11 at % Zn	2·643	48·7	48·2

centres of distortion retain their individuality so that the hardness is affected additively. We shall not of course expect such a simple relation when the concentration of a ternary solid solution becomes great enough for the interaction between the two solutes to be appreciable.

From figure 3 and table 3 it will be seen that the increase in hardness for a given lattice distortion is smaller for the solutes cadmium, indium, tin, and antimony which follow silver in the Periodic Table than for solutes from other periods. This suggests that a given mean lattice distortion is less severe from the point of view of hardening when the solvent and solute have the same ionic structure and the same quantum number of their valency electrons, than when solvent and solute are farther apart in the Periodic Table. This conclusion is confirmed by the work of Norbury, whose results showed that at a given mean lattice distortion the hardening effect is much less for zinc, which follows copper in the Periodic Table, than for the other solutes examined, which were from different periods.

In comparing the hardness of the silver and copper alloys, the systems copper-zinc and silver-cadmium appear strictly analogous, and it is of interest to note that if the P_u values are plotted against the lattice distortions, the initial gradient

of the curve is greater for the silver-cadmium alloys. Silver is more compressible than copper, and the above facts suggest, therefore, that when the solid solutions are analogous, a given increase in the mean lattice spacing produces a greater hardening effect when the solvent metal is more compressible. This is in agreement with recent views of G. V. Raynor (1941) on age-hardening, and suggests that, other things being equal, a high compressibility results in a disturbance being spread over a wider region, and so involves a greater interference with the deformation process.

The above discussion refers exclusively to the P_u values, but the present work also throws some light on the rate of hardening of solid solutions with increasing
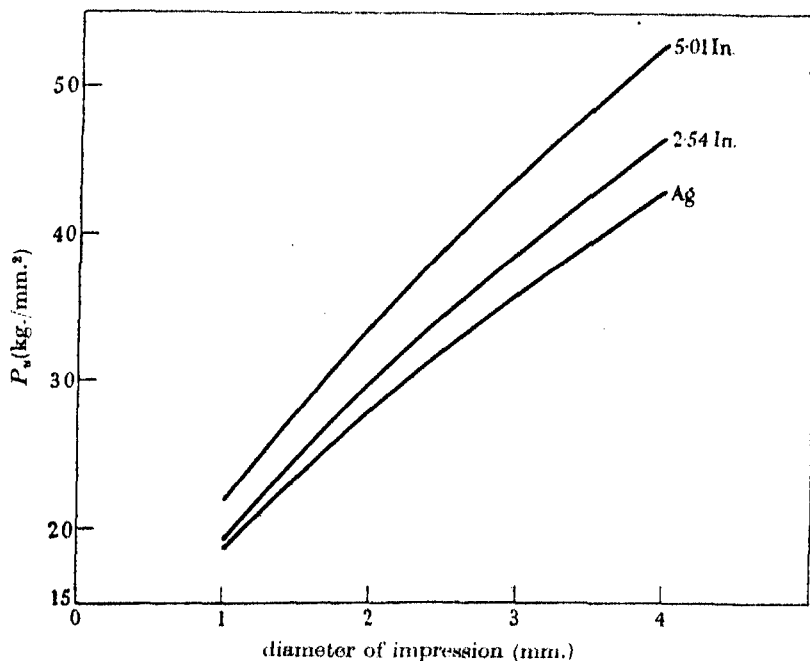
FIGURE 4

deformation. Reference to the tables will show that the P_u values are always greater than the P_a values, so that the Meyer hardness increases with increasing load. Figure 4 shows the variation of Meyer hardness with increasing diameter of the impression from 1 to 4 mm., the curves referring to pure silver, and to the two alloys with indium. It will be seen that the rate of increase in hardness with increasing diameter of the impression is greater for the two alloys than for the pure metal, and this is a general principle. In the whole series of alloys, no case was found in which the gradient of the curve was less for the alloy than for pure silver, and although in the cadmium alloys the two curves were almost parallel, it seems reasonable to conclude that the formation of a solid solution increases the rate of hardening accompanying increased deformation.

The authors must express their gratitude to Professor C. N. Hinshelwood, F.R.S., for laboratory accommodation, and many other facilities which have greatly assisted the present research. Grateful acknowledgement of financial assistance is made to the Council of the Royal Society, the British Non-Ferrous Metals Research Association, and the Curators of the Leigh Fund of the University of Oxford, and to the Imperial Chemical Industries, Ltd., for the loan of apparatus.

REFERENCES

Angus, H. & Summers, P. 1925 *J. Inst. Met.* **33**, 115.
Hume-Rothery, W., Lewin, G. F. & Reynolds, P. W. 1936 *Proc. Roy. Soc.* A, **157**, 167.
Meyer, E. 1908 *Z. Ver. dtsch. Ing.* **52**, 645.
Norbury, A. L. 1923 *J. Inst. Met.* **29**, 423.
Norbury, A. L. 1923 *Trans. Faraday Soc.* **19**, 586.
O'Neill, H. 1934 *The hardness of metals and its measurement.* London.
Owen, E. A. & Roberts, E. W. 1939 *Phil. Mag.* **26**, 294. Also Owen, E. A. & Edmunds, I. G. 1938 *J. Inst. Met.* **63**, 265.
Raynor, G. V. Unpublished work.

The elastic scattering of fast positrons by heavy nuclei

By H. S. W. MASSEY, F.R.S.

The angular distribution of fast positrons scattered elastically by mercury nuclei is investigated in detail using Dirac's equations in conjunction with numerical tables provided by Bartlett & Watson for the corresponding fast electrons. The ratio of the scattered intensity to that given in the Rutherford formula is obtained as a function of angle for positrons with energies between 25,000 and 1.7×10^6 eV. In all cases the ratio is less than unity, decreases with increasing angle of scattering and, in contrast to electrons, does not depend very markedly on the nuclear charge. The maximum asymmetry produced by double scattering of positrons at 90° from mercury nuclei is less than 1 %.

It is suggested that experimental investigation of the scattering of fast positrons would be useful in assisting to resolve the present unsatisfactory situation in which there are a number of outstanding divergences between experimental and theoretical results relating to the scattering of fast electrons.

The theory of the scattering of fast electrons by the Coulomb fields of atomic nuclei was first developed by Mott (1929, 1932). Using Dirac's equation, he derived an explicit formula for the scattered intensity, valid for light nuclei of charge Ze such that $2\pi Ze^2/hc \ll 1$. No corresponding formula can be obtained for scattering by heavy nuclei, but Mott carried out some numerical calculations for gold. In particular he determined the extent to which an electron beam is polarized by double scattering at 90° by gold atoms. These numerical calculations have recently been greatly extended by Bartlett & Watson (1940) for scattering by mercury nuclei. They derived in detail the angular distributions of the scattered electrons for electron energies ranging from 20,000 to 1.7×10^6 eV. Very marked deviations of these distributions from the forms to be expected in scattering by light nuclei, clearly

capable of experimental observation, were revealed. Cloud-chamber experiments have been carried out in mercury vapour by Barber & Champion (1938) using the β-particles from radium E (0·6–1·1 MeV energy). The angular distributions which they observed do agree in form with those calculated by Bartlett & Watson, but the theoretical intensity of scattering is over 6 times that observed. This unsatisfactory result is by no means confined to this case. The considerable polarization (up to 16 %) on double scattering, predicted by Mott for gold and confirmed by Bartlett & Watson's calculations for mercury, has never been observed, and conflicting results have been obtained in single scattering experiments in other gases and from foils. Thus, while Oppenheimer & Fowler (1938) obtained agreement with theory for the scattering of 5–17 MeV electrons by lead, Stepanowa (1937) found up to 30 times the theoretical scattered intensity for 1·5–3·0 MeV electrons scattered in nitrogen. Again, Klarmann & Bothe (1936) found only one-fifth of the theoretical scattering for 0·5–2·5 MeV electrons in xenon, in contrast to Sen Gupta (1939) who found agreement with theory for 2 MeV electrons in the same gas. Similar marked discrepancies exist in inelastic collision phenomena involving electrons in this energy range, i.e. in the probability of emission of radiation and of pair formation. A summary of these has been given by Champion (1939).*

In view of this position it becomes of special interest to investigate, both theoretically and experimentally, the scattering of positrons. For very light nuclei this should differ only slightly from that of the corresponding electrons, but for heavy nuclei the change in sign of the Coulomb force can be expected, through the medium of spin-orbital interaction, to affect profoundly the form of the theoretical angular distributions. This is confirmed by a detailed calculation for mercury nuclei, the results of which we describe in this paper. It is clear from these results that experiments on positron scattering, which should not prove difficult now that suitable positron sources are available, are likely to clarify the position considerably. For example, if it proves that the shapes of the observed and calculated angular distributions for positrons scattered by mercury agree in the same way as do those for electron scattering, there will be strong evidence in favour of the validity of the theory for both particles in spite of any failure of the theory to reproduce the observed intensity—the presumption will be that the observed rather than the calculated intensities are at fault.

In deriving the theoretical results we have assumed that the positron behaves, in the scattering phenomena, as a Dirac particle of positive charge, so the validity of this assumption can also be checked by comparison with observed results.

FORMULAE AND METHOD OF COMPUTATION

Consider an unpolarized beam of positrons, of mass m and uniform velocity v, incident on a scatterer of nuclear charge Ze. The differential cross-section for

---

* In a recent paper, Brailovsky & Leipunsky (1941) have reported experiments on the scattering of electrons with energies 1·5–3 MeV in argon and they find good agreement with Mott's theory.

scattering in directions making angles between θ and $\theta + d\theta$ with the direction of incidence will be given by

$$2\pi I(\theta)\sin\theta\, d\theta = 2\pi k^{-2}\{|f|^2 + |g|^2\}\sin\theta\, d\theta, \qquad (1)$$

where f, g differ from the corresponding quantities derived by Mott (1932) for electron scattering only in the replacement of $+Z$ by $-Z$. Thus we have

$$f = iq'F + G, \qquad (2)$$

$$g = [-iq'(1+\cos\theta)F + (1-\cos\theta)G]/\sin\theta, \qquad (3)$$

where

$$F = -\tfrac{1}{2}i\Sigma[lS_{\rho_l} + (l+1)S_{\rho_{l+1}}](-1)^l P_l(\cos\theta), \qquad (4)$$

$$G = -\tfrac{1}{2}i\Sigma[l^2 S_{\rho_l} - (l+1)^2 S_{\rho_{l+1}}](-1)^l P_l(\cos\theta), \qquad (5)$$

$$S_\rho = \frac{e^{-\pi i\rho}}{\rho - iq}\frac{\Gamma(\rho+iq)}{\Gamma(\rho-iq)}, \quad \rho_l = (l^2-\alpha^2)^{\frac{1}{2}}, \quad q = \alpha/\beta,$$

$$\beta = v/c, \quad q' = q(1-\beta^2)^{\frac{1}{2}}, \quad \alpha = 2\pi Ze^2/hc, \quad k = 2\pi mv(1-\beta^2)^{-\frac{1}{2}}/h.$$

When $\alpha \ll 1$ we have the approximate formula derived by Mott (1929)

$$|f|^2 + |g|^2 = \frac{q^2}{4}\operatorname{cosec}^4\tfrac{1}{2}\theta[1 - \beta^2\sin^2\tfrac{1}{2}\theta - \pi\alpha\beta\cos^2\tfrac{1}{2}\theta\sin\tfrac{1}{2}\theta + \text{terms of order } \alpha^2]. \qquad (6)$$

For small angles of scattering Bartlett & Watson derive the formula

$$|f|^2 + |g|^2 = \frac{q^2}{4}\cot^4\tfrac{1}{2}\theta\sec^2\tfrac{1}{2}\theta[1 - \pi\alpha\beta\sin\tfrac{1}{2}\theta\cos\chi], \qquad (7)$$

where

$$e^{i\chi} = \Gamma(\tfrac{1}{2}+iq)\,\Gamma(1-iq)/\{\Gamma(\tfrac{1}{2}-iq)\,\Gamma(1+iq)\}. \qquad (8)$$

This agrees with (6) when $q \ll \tfrac{1}{2}$ and θ is small.

When α and θ are not small, numerical methods must be applied to the formulae (4) and (5). Owing, however, to the very slow convergence these series cannot be summed directly. As shown by Mott (1932) and by Bartlett & Watson (1940) it is possible to write

$$F = F_0 + F_1, \quad G = G_0 + G_1 + G_2 + G_3,$$

where $F_0 = \tfrac{1}{2}i(\sin\tfrac{1}{2}\theta)^{-2iq}\,\Gamma(1+iq)/\Gamma(1-iq),$

$\quad F_1 = \tfrac{1}{2}i\Sigma[lD_l + (l+1)D_{l+1}](-1)^l P_l(\cos\theta),$

$\quad G_0 = iqF_0\cot^2\tfrac{1}{2}\theta,$

$\quad G_1 = (\pi\alpha^2/8)(1+\cos\theta)(2\sin\tfrac{1}{2}\theta)^{-2iq-1}F(\tfrac{1}{2}, -iq+1, 2; \cos^2\tfrac{1}{2}\theta)$
$$\qquad\qquad\qquad \times \Gamma(-iq+\text{L})\,\Gamma(-iq-1)\{\Gamma(1-2iq)\}^{-1},$$

$\quad G_2 = (-iF_0/8q)[\pi^2\alpha^4 - 4\alpha^2(1-2iq)],$

$\quad G_3 = \tfrac{1}{2}i\Sigma[E_l - E_{l+1}](-1)^l P_l(\cos\theta),$

$\quad D_l = S_l - S_{\rho_l},$

$\quad E_l = l^2 D_l + (\pi i\alpha^2/2)\,lS_l - \tfrac{1}{8}[\pi^2\alpha^4 - 4\alpha^2(1-2iq)]S_l.$

The series F_1, G_3 are readily summed as the terms converge rapidly. Similarly G_1 can be easily calculated for all θ if use is made of the connexion formula for hypergeometric functions. As F_0, G_0 and G_2 are in closed form, the numerical evaluation of $I(\theta)$ can then be carried out without difficulty.

The work is greatly simplified if use is made of the extensive tables given by Bartlett & Watson (1940) for electron scattering. If we distinguish by superscripts $+$, $-$ the corresponding quantities for positrons and electrons respectively, we have

$$F_0^+ = -\bar{F}_0^-, \quad G_0^+ = -\bar{G}_0^-, \quad G_1^+ = \bar{G}_1^-, \quad G_2^+ = -\bar{G}_2^-.$$

As F_0^-, G_0^-, G_1^-, G_2^- are tabulated by Bartlett & Watson, it is only necessary to compute F_1 and G_3 specially for the positron case. Even for these, further use may be made of Bartlett & Watson's tables of ρ_l, $S_{\rho_l}^-$, S_l^- and E_l^-, for we have

$$S_{\rho_l}^+ = \exp(-2\pi i \rho_l)\bar{S}_{\rho_l}^-, \quad S_l^+ = \bar{S}_l^-,$$

so $$D_l^+ = \exp(-2\pi i \rho_l)S_{\rho_l}^- - \bar{S}_l^-, \quad E_l^+ = \bar{E}_l^- - l^2(S_{\rho_l}^+ - \bar{S}_{\rho_l}^-) + \pi i \alpha^2 l \bar{S}_l^-.$$

The percentage asymmetry to be expected on double scattering of the beam at $90°$ is given by 200δ, where

$$\delta = \{|\,f\bar{g} - g\bar{f}\,|/(|\,f\,|^2 + |\,g\,|^2)\}^2,$$

f, g being calculated for $\theta = 90°$.

RESULTS AND DISCUSSION

In figure 1 the angular distributions of positrons scattered by mercury nuclei are illustrated for a variety of positron energies. These distributions are represented as the ratio $I(\theta)/R$, where $R = q^2 \operatorname{cosec}^4 \tfrac{1}{2}\theta/k^2$ is the angular distribution function given by the ordinary Rutherford formula with allowance only for the variation of mass with velocity.

In contrast to electrons (Bartlett & Watson 1940) the intensity of scattered positrons is always less than that given by the Rutherford distribution function R. Nevertheless the deviations from that formula are sufficiently marked to be observed easily experimentally.

Figure 2 illustrates a comparison of various distribution functions $I(\theta)/R$ for electrons and for positrons of $1\cdot7$ MeV energy, calculated by accurate theory and by use of certain approximations. It is clear that there is a very marked difference in the distributions for the two particles. Further, the positron distribution does not differ very markedly from that given by Mott's formula (6) with $\alpha = 0$, despite the fact that the term $\pi\alpha\beta \cos^2 \tfrac{1}{2}\theta \sin \tfrac{1}{2}\theta$ is larger than the middle term except for large θ. This is probably due to the series, of which $\pi\alpha\beta \cos^2 \tfrac{1}{2}\theta \sin \tfrac{1}{2}\theta$ is the first term, being an alternating one for positrons but not for electrons. The formula (7) tends to underestimate the positron scattering.
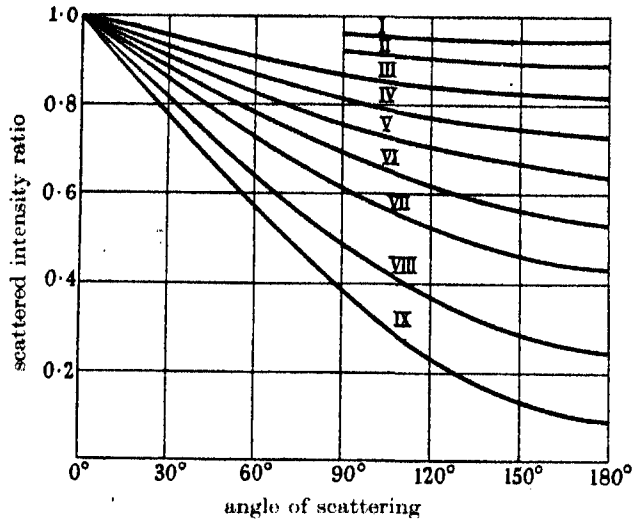
FIGURE 1. Angular distributions of positrons of various energies scattered by mercury nuclei. The scattered intensity is given as the ratio to that, $q^2 \operatorname{cosec}^4 \frac{1}{2}\theta/k^2$, given by the Rutherford formula. Curves I–IX respectively correspond to positrons of energies 0·046, 0·086, 0·145, 0·232, 0·314, 0·463, 0·666, 1·28 and 3·35 mc^2.
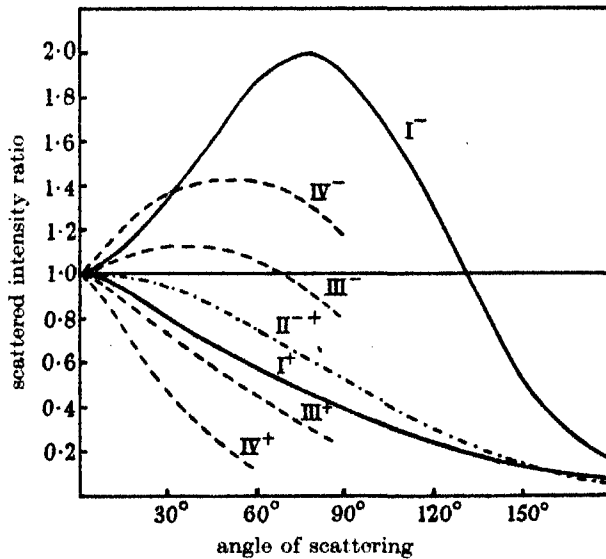
FIGURE 2. Comparison of angular distributions (expressed as ratios of the Rutherford distribution) for electrons and for positrons, of energy 3·35 mc^2, scattered by mercury nuclei. Curves I calculated without approximation. Curve II calculated from formula (6), without inclusion of terms in α. Curves III calculated from Bartlett & Watson's 'small-angle' formula (7). Curves IV calculated from Mott's 'light element' formula (6). Superscripts −, + respectively distinguish electrons and positrons.

Although it is not possible in practice to investigate the polarization produced by double scattering of positrons, it is of interest to note the contrast between the theoretical asymmetries for positrons and for electrons. This is illustrated in figure 3 in which our results for positrons as well as those computed by Bartlett & Watson (1940) for electrons are given. It will be seen that very little polarization is effected in a positron beam of any energy, compared with the possible maximum of 13 % for 120,000 eV electrons. This is because the repulsive field of the nucleus largely prevents close approach of the positrons so they do not come under the action of a force large enough to affect their spin orientation. On the other hand, electrons scattered by heavy nuclei do come under the action of a large force, there being a strong nuclear attraction tending to concentrate their density in the neighbourhood of the nucleus.
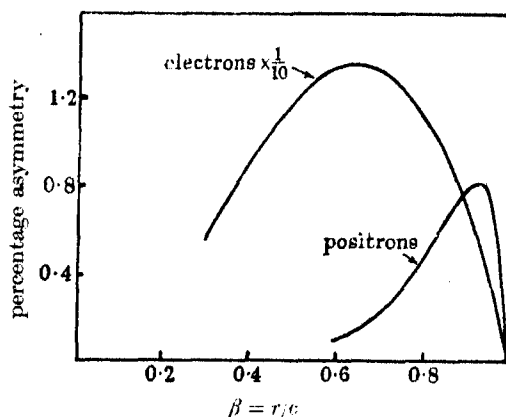
FIGURE 3. Percentage asymmetry in double scattering at 90° by mercury nuclei, calculated for electrons (Bartlett & Watson 1940) and for positrons.

The calculations of this paper were greatly facilitated by the availability of a copy of Bartlett & Watson's tables for electron scattering, kindly sent to the author by Dr Bartlett.

REFERENCES

Barber & Champion 1938 *Proc. Roy. Soc.* A, **168**, 159.
Bartlett & Watson 1940 *Proc. Amer. Acad. Arts Sci.* **74**, 53.
Brailovsky & Leipunsky 1941 *J. Phys. U.S.S.R.* **4**, 485.
Champion 1939 *Rep. Prog. Phys.* **5**, 348.
Klarmann & Bothe 1936 *Z. Phys.* **101**, 489.
Mott 1929 *Proc. Roy. Soc.* A, **124**, 426.
Mott 1932 *Proc. Roy. Soc.* A, **135**, 429.
Oppenheimer & Fowler 1938 *Phys. Rev.* **53**, 928.
Sen Gupta 1939 *Proc. Phys. Soc.* **51**, 355.
Stepanowa 1937 *Phys. Z. Sowjet.* **12**, 550.

A derivation of the tidal equations

By Harold Jeffreys, F.R.S.

As the validity of Laplace's tidal equations is under discussion (Proudman 1942), I think it may be worth while to deal with a difficulty in their derivation pointed out once by Mr E. Gold (1926) in a discussion at the Royal Meteorological Society. It concerns the use of z, the normal distance from a standard level surface, as one of the position co-ordinates. The other level surfaces are approximately surfaces where gz is constant; but g varies by a factor of about 1/200 from pole to equator. Hence in an ocean 4 km. deep, if we take the standard level as mean sea-level, z on the level surfaces near the bottom will vary by about 20 m., which is much more than the height of the equilibrium tide. We need some explanation of why differences of pressure for constant z can be ignored, especially since the usual form of the equations suggests that the condition for equilibrium is that the pressure is constant over surfaces of constant z. Similar considerations apply to atmospheric motions.

It is clear from hydrostatics that equilibrium can exist if the pressure and density are constant over the level surfaces, and this would be true even if we took into account the small departures of the level surfaces from symmetry about the axis. The difficulty arises entirely from the fact that when the ellipticity is taken into account the three directions of increasing θ (co-latitude), ϕ (longitude) and z (height from a given surface) are not strictly orthogonal; when we are above or below the standard surface the direction of the normal to the standard surface from the point is not at right angles to the direction of the meridian in the level surface, and if we use z as a co-ordinate this difference should be allowed for.

On the other hand, we really have a possible set of exactly orthogonal co-ordinates, and it seems a pity not to use them. Let U be the symmetrical part of the earth's gravitational potential (i.e. its average with regard to longitude) and ϖ the distance of a point from the axis of rotation. Then the geopotential, with the sign changed, is

$$\Psi = -U - \tfrac{1}{2}\varpi^2\omega^2, \tag{1}$$

and is constant over a symmetrized level surface. Let θ be the angle between the normal to the level surface at the point considered and the polar axis (the geographic co-latitude) and ϕ the east longitude. Then the directions of Ψ, θ, ϕ increasing are exactly orthogonal and are a right-handed set of axes. The elements of length for small changes in Ψ, θ, ϕ are $d\Psi/g$, $R\,d\theta$, $\varpi\,d\phi$, where g is local gravity and R the radius of curvature of the meridian. Let the work per unit mass due to forces other than those expressed in U, done on a particle in a small displacement, be $J\,d\Psi + \Theta\,d\theta + \Phi\,d\phi$. The kinetic energy per unit mass is

$$2T = R^2\dot\theta^2 + \varpi^2(\omega + \dot\phi)^2 + \dot\Psi^2/g^2, \tag{2}$$

and Lagrange's equations give

$$\frac{d}{dt}(R^2\dot\theta) - \tfrac{1}{2}\dot\theta^2\frac{\partial R^2}{\partial\theta} - \tfrac{1}{2}(\omega+\dot\phi)^2\frac{\partial\varpi^2}{\partial\theta} - \tfrac{1}{2}\dot\Psi^2\frac{\partial}{\partial\theta}\left(\frac{1}{g^2}\right) = \frac{\partial U}{\partial\theta} + \Theta,$$

$$\frac{d}{dt}\{\varpi^2(\omega+\dot\phi)\} = \Phi, \qquad\qquad (3)$$

$$\frac{d}{dt}\left(\frac{1}{g^2}\dot\Psi\right) - \tfrac{1}{2}\dot\theta^2\frac{\partial R^2}{\partial\Psi} - \tfrac{1}{2}(\omega+\dot\phi)^2\frac{\partial\varpi^2}{\partial\Psi} - \tfrac{1}{2}\dot\Psi^2\frac{\partial}{\partial\Psi}\left(\frac{1}{g^2}\right) = \frac{\partial U}{\partial\Psi} + J.$$

Neglect squares of the velocities and eliminate $\dot\theta$, $\dot\phi$, $\dot\Psi$ in favour of the velocity components

$$(u, v, w) = (R\dot\theta, \varpi\dot\phi, \dot\Psi/g). \qquad\qquad (4)$$

$$\frac{\partial}{\partial t}(Ru) - 2\omega\frac{\partial\varpi}{\partial\theta}v = \frac{\partial U}{\partial\theta} + \tfrac{1}{2}\omega^2\frac{\partial\varpi^2}{\partial\theta} + \Theta = -\frac{\partial\Psi}{\partial\theta} + \Theta = \Theta,$$

$$\frac{\partial}{\partial t}(\varpi v) + \frac{d}{dt}(\varpi^2\omega) = \Phi, \qquad\qquad (5)$$

$$\frac{d}{dt}\left(\frac{w}{g}\right) - 2\omega\frac{\partial\varpi}{\partial\Psi}v = \frac{\partial U}{\partial\Psi} + \tfrac{1}{2}\omega^2\frac{\partial\varpi^2}{\partial\Psi} + J = -1 + J,$$

and

$$\frac{d}{dt}(\varpi^2\omega) = 2\varpi\omega\left(\frac{u}{R}\frac{\partial\varpi}{\partial\theta} + wg\frac{\partial\varpi}{\partial\Psi}\right) = 2\varpi\omega(u\cos\theta + w\sin\theta). \qquad (6)$$

In a fluid of density ρ we have

$$(\Theta,\Phi,J) = -\frac{1}{\rho}\left(\frac{\partial p}{\partial\theta}, \frac{\partial p}{\partial\phi}, \frac{\partial p}{\partial\Psi}\right) + \left(\frac{\partial U_1}{\partial\theta}, \frac{\partial U_1}{\partial\phi}, \frac{\partial U_1}{\partial\Psi}\right), \qquad (7)$$

where U_1 is the potential of any gravitational forces not included in U. Also if r is the distance from the centre

$$\varpi = r\{1 + O(e)\}\sin\theta; \quad R = r + O(er), \qquad\qquad (8)$$

and if we neglect products of the velocities with the ellipticity we can make the substitutions

$$\frac{\partial\varpi}{\partial\theta} = r\cos\theta; \quad \frac{\partial\varpi}{\partial\Psi} = \frac{r}{g}\sin\theta. \qquad\qquad (9)$$

Then

$$\frac{\partial u}{\partial t} - 2\omega\cos\theta v = -\frac{\partial p}{\rho r\,\partial\theta} + \frac{\partial U_1}{r\,\partial\theta},$$

$$\frac{\partial v}{\partial t} + 2\omega\cos\theta u + 2\omega\sin\theta w = -\frac{\partial p}{\rho r\sin\theta\,\partial\phi} + \frac{\partial U_1}{r\sin\theta\,\partial\phi}, \qquad (10)$$

$$\frac{\partial w}{\partial t} - 2\omega\sin\theta v = -g\left(1 + \frac{\partial p}{\rho\,\partial\Psi}\right) + \frac{g\partial U_1}{\partial\Psi}.$$

The left sides of these equations are in the usual form. The explicit appearance of r in the first two on the right makes them applicable to a deep layer as well as one whose depth is small compared with the radius; they could be applied, for instance, to motions in the gaseous parts of Jupiter and Saturn. If

$$p = -\int \rho d(\Psi' - U_1), \tag{11}$$

and ρ is a function of p only, then the equations are satisfied with $(u, v, w) = 0$ for all time; this will be possible if U_1 is independent of t. Hence any non-periodic part of the gravitational potential, such as the asymmetrical part of the earth's field, produces no currents; it only makes p constant over the level surfaces with this part of the potential included, and need not be considered in tidal problems, as indeed we should expect from hydrostatic theory.

The advantages of the present derivation are that (1) it makes it clear why the terms in ω^2 cancel exactly, (2) it shows that all neglected terms are of the order of the ellipticity times those retained, (3) the velocity components u, v, w are strictly perpendicular.

The equation of continuity needs no special comment. In a deep layer it has the usual form for three-dimensional orthogonal co-ordinates; in a shallow one the form given by Lamb is correct when the 'long-wave' approximation applies. This approximation is fully discussed by Proudman.

REFERENCES

Gold, E. 1926 *Quart. J.R. Met. Soc.* **52**, 102.
Proudman, J. 1942 *Proc. Roy. Soc.* A, **179**, 261–288.

The influence of temperature on boundary lubrication

white oil, liquid alcohols, and the lower fatty acids the motion proceeds in a series of jerks: when the higher fatty acids are used as lubricants the motion is smooth.

Tabor (1940) found that in the case of steel surfaces lubricated with some commercial oils and fatty acids which gave smooth sliding at room temperature, the motion changed to the jerky 'stick-slip' type as the surface was heated. This transition from smooth sliding to stick-slip is reversible with temperature and characteristic of the lubricant employed. Tabor suggested that the transition is due to a disorientation or desorption of the adsorbed film, and the temperature at which it occurs is a measure of the strength with which the film is adsorbed on the surface.

This paper describes an investigation of this effect for a number of pure chemical compounds used as lubricants for mild steel and stainless steel surfaces, both as an excess and as built-up films of known molecular thickness.

The experiments with excess of lubricant on hydrocarbons, alcohols, and two of the fatty acids were made in collaboration with Tabor, who set up the apparatus used in this investigation.*

EXPERIMENTAL

The apparatus employed in this investigation was similar to that described by Bowden & Leben (1939).

It consisted of a flat plate with its face horizontal, carried on a horizontal slide driven slowly and uniformly by a hydraulic piston. An approximately hemispherical contact was pressed vertically against the face of the plate, with a known pressure, by an adjustable spring.

The friction acting on the contact was measured by the deflexion of the suspension which carried it. In order to record rapid fluctuations in friction a system with a rapid response—and therefore a stiff restraint and low moment of inertia—was required. This was achieved by using two vertical piano wires as a bifilar suspension for a bar of Electron which carried the contact. The bar was horizontal, perpendicular to the direction of motion, and above the face of the plate. The effective length of the piano wires was 6 cm. and the wires were 3 cm. apart. One end of the Electron bar was clamped to the centre of one of the wires, the other wire being clamped near the centre of the bar and 2·5 cm. from the contact which was carried at the other end of the bar and projected over the plate. The friction acting on the contact rotated the bar against the restoring force of the piano-wire suspension, the rotation being recorded by a galvanometer mirror attached to the arm.

For the range of loads 2–10 kg., 1·22 mm. diameter piano wire was used, and for loads of ½–3 kg. 0·38 mm. diameter piano wire was employed. With the thick

wires the natural frequency of vibration of the system recording the friction was of the order of 1000, which was about ten times greater than with the thin wires.

A heating element built into the apparatus beneath the plate enabled the latter to be heated electrically, the rate of heating being controlled by rheostats in the circuit. The temperature of the surface of the plate in the neighbourhood of the contact was measured by a thermocouple consisting of a copper point and a constantan point resting on the surface and connected in series with a mirror galvanometer.

The lower surface was prepared by grinding on a flat wet lead lap with 320 carborundum, cleaned with rouge and water, washed in 20 % caustic soda solution and rewashed with water.

The plate was then placed in the vapour of boiling acetone to remove the water. A bright reflecting surface crossed by numerous fine scratches was obtained. The hemispherical contact was polished on 0000 emery paper. The plate and the contact were made from the same piece of mild steel of the following composition: carbon 0·13 %, silicon 0·21 %, manganese 0·59 %, sulphur 0·018 %, phosphorus 0·015 %, chromium 0·05 %, nickel 3·42 %. The steel was taken from the bottom of an ingot to ensure uniformity.

In the experiments in which an excess of lubricant was employed approximately 0·1 g. of the lubricant was placed on the surface, which, in the case of solids, was warmed to melt the lubricant. The position of the contact was then adjusted so as to come within the drop of lubricant. An excess of the lubricant was then maintained round the contact by surface tension.

In carrying out a determination of the transition temperature the plate was set in motion and slowly heated at a uniform rate. The friction was recorded on a moving photographic film by a beam of light, reflected from the mirror attached to the Electron bar carrying the contact, and focused on the slit of a camera. A simultaneous record of the temperature of the surface of the plate was obtained by another beam of light reflected from the mirror galvanometer and also focused on the camera.

THE EFFECT OF THE LOAD AND CHARACTERISTICS OF THE RECORDER ON THE FRICTIONAL BEHAVIOUR

The mean values of a large number of determinations, with excess of oleic acid as lubricant, of the coefficient of friction in the region of smooth sliding, and the temperature at which the transition from smooth sliding to stick-slip occurs, for loads of ½–10 kg., are plotted in figure 1.

As the load used in experiments was increased, both the coefficient of friction and the transition temperature increased until constant values were reached at loads of 3 kg. and over. The question arose whether the differences in coefficient of friction and transition temperature between runs at high and low loads were due to some mechanical effect (such as the contact gouging a track in the plate at high

loads) or to a variation in the thickness of the lubricating film with load. To test this point, experiments were made in which the amount of lubricant present was limited to a unimolecular film on each of the test-pieces, and it was then found that the coefficient of friction and the transition temperature had the same values for loads of $\frac{1}{2}$ and 4 kg. These experiments were carried out with ethyl stearate and elaidic acid as lubricants, the unimolecular films being deposited from water on the test-pieces by the Langmuir-Blodgett technique. Stainless steel test-pieces were used for this purpose.
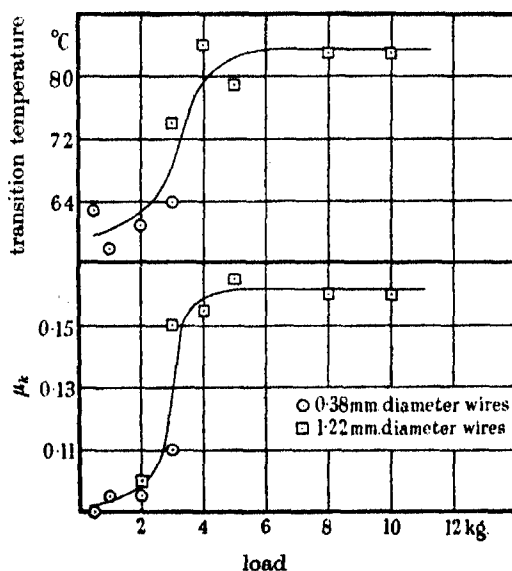
FIGURE 1. Oleic acid.

This result shows that the dependence of the coefficient of friction on load, observed with excess of oleic acid, was due to a variation of the thickness of the lubricating film, with load. The limiting high value at high loads is roughly the same as that obtained in the experiments on unimolecular films of methyl stearate, ethyl stearate, elaidic acid, stearic acid and myristone (described in a later section of this paper), all of which have molecules of approximately the same length as oleic acid.

The hemispherical contact becomes deformed under the applied load (Bowden & Tabor 1939), the area of contact increasing with the load. The pressure distribution over the area of contact will not be uniform: in the centre where it is highest it may be sufficient even at comparatively low loads to penetrate to the primary film. At high loads penetration to the primary film will occur over a greater proportion of the area of contact. At low loads the frictional behaviour is mainly due to the contact sliding over the outer oriented layers, while above 3 kg. the

frictional behaviour is predominantly affected by the primary adsorbed layer. Further support for this view is afforded by experiments with ethyl stearate as lubricant. When an excess of this substance was present no stick-slips occurred at room temperature and the coefficient of friction remained constant at 0·18 over the range of loads of 2–10 kg. using the high-frequency friction recorded (1·22 mm. diameter wires), but at lower loads with the low-frequency recorder (0·38 mm. diameter wires) stick-slips occurred at room temperature and increased in size with rise of temperature. When a unimolecular film was present a steady value of 0·18 was obtained for the coefficient of friction at a load of ½ kg.; when a film three molecules thick was used stick-slips occurred. Evidently at the low loads used with the low-frequency recorder the frictional behaviour is influenced by layers beyond the primary adsorbed film. The occurrence of stick-slips in this case is not due to the disorientation of the primary film, which is shown elsewhere in this paper to be responsible for the occurrence of stick-slips at higher loads.

These results show that orientation in the case of oleic acid extends into the liquid beyond the primary layer adsorbed on the steel surface. In this connexion it may be noted that Trillat (1929, 1930) has shown that orientation of long-chain acids may extend some little distance from a surface into the interior of a liquid, and Andrews's electron-diffraction experiments (1936) indicate orientation extending beyond the primary layer.

These results clearly demonstrate the importance of using high loads if it is desired to investigate lubrication under true boundary conditions. A load of 4 kg. on the hemispherical contact and a sliding speed of 0·005 cm./sec. were used in all the following experiments.

I. EXPERIMENTS WITH AN EXCESS OF LUBRICANTS ON MILD STEEL SURFACES

Hydrocarbons

Hexadecane, diphenylmethane, paraffin wax, diphenyl and bis-*p*-diphenyl-methane all gave smooth sliding in the solid state and a sharply defined transition to stick-slips at the bulk melting-point. The photographic record of the friction obtained in the case of diphenylmethane is reproduced in figure 2. When melting was observed on the surface in the neighbourhood of the contact, the thermocouple circuit was momentarily broken. This point is recorded by a deflexion in the temperature trace.

Alcohols

Lauryl and ceryl alcohol gave smooth sliding in the solid state and a transition to stick-slips at the melting-point.

Ketones

Methyl hepta-decyl ketone, myristone and benzophenone gave smooth sliding in the solid state and a transition to stick-slips at the melting-point.

Amides

Acetamide, pelargonamide and myristamide gave smooth sliding in the solid state and a transition to stick-slips at the melting-point.
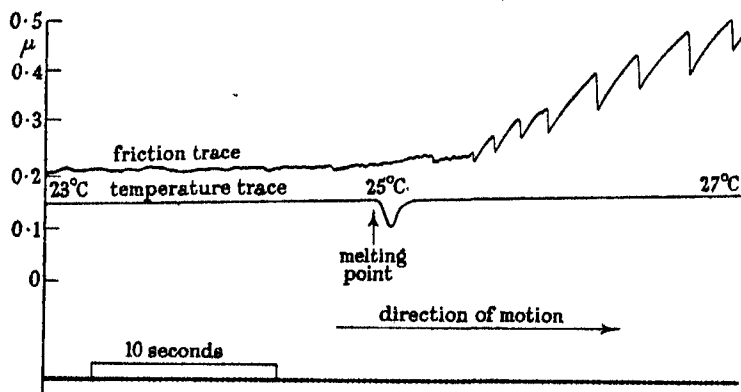
FIGURE 2. Diphenylmethane.

Acids

No change in the frictional behaviour occurred as the fatty acids were heated through their melting-points. Figure 3 shows the friction trace of elaidic acid as it passed through its melting-point. At a temperature considerably above the melting-point the motion changed from smooth sliding to stick-slips as shown in
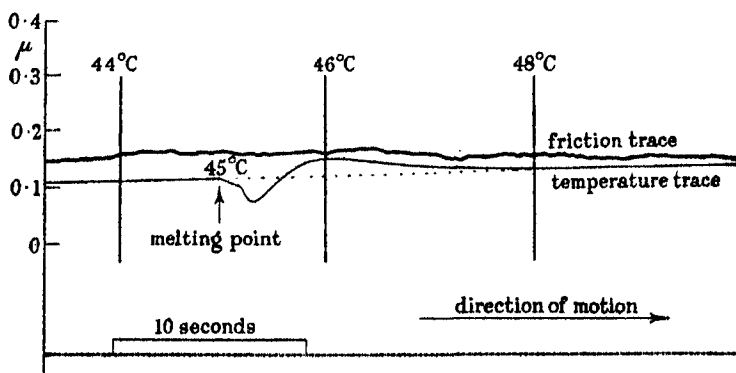
FIGURE 3. Elaidic acid.

figure 4, which is typical of the results obtained with the fatty acids. When the transition occurs at a temperature considerably above the melting-point, it is not so sharply defined as when it occurs at or near the melting-point, but extends over a range of about 5° C. The value of the transition temperature also varies in different parts of the surface due to the difficulty of obtaining a surface of uniform

finish and uniform activity. Successive values obtained in the case of pelargonic acid were: 95, 90, 95, 95, 100, 90, 85, 80, 85, 80, 90° C, giving a mean value of 90° C (to the nearest 5° C).

A number of the fatty acids of the series C_nH_{2n+1}.COOH from butyric ($n = 3$) to stearic ($n = 17$) were examined. The transition temperature could not be determined in the case of the lower members of the series owing to their volatility, and the rapidity with which they attacked the surface forming a layer of the ferric salt on which smooth sliding persisted to high temperatures.
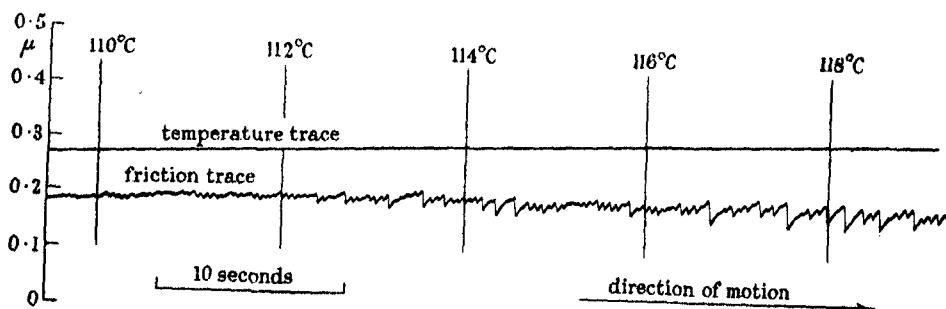
FIGURE 4. Elaidic acid.

In the case of stearic acid the transition temperature decreased with successive determinations, eventually reaching a constant value of 110° C. When this stage was reached the drop of stearic acid on the surface was observed to be slightly brown: it was then shown by a chemical test to contain iron. Some experiments were made with pure dry ferric stearate, which gave a transition at 115° C. According to Lawrence (1937) the true melting-point of ferric stearate is 113° C. In the case of ferric stearate the transition occurs approximately at the melting-point. These results show that the initial high value of the transition temperature of stearic acid is the true value of the acid, the subsequent lower values being due to the formation of the ferric salt. Palmitic acid behaved in a similar manner, but the other acids, intermediate between palmitic and capric, gave a series of values constant to within 5° C. Although it is well known that the acidic properties of this series of acids becomes more marked as the length of the chain decreases, the transition temperature, on the other hand, like the melting-point, rises as the chain length increases. The rate of attack of the acid on the steel surface increases rapidly with rise of temperature, so that, although the higher members of the series may not show such marked acidic properties as the lower members, they may attack the steel more rapidly at their higher transition temperature.

The results obtained with this series of fatty acids are shown in figure 5, from which it will be seen that the difference between the transition temperature and the melting-point is approximately 70° C for the acids in which the length of the hydrocarbon chain varies from 8 to 17 carbon atoms.
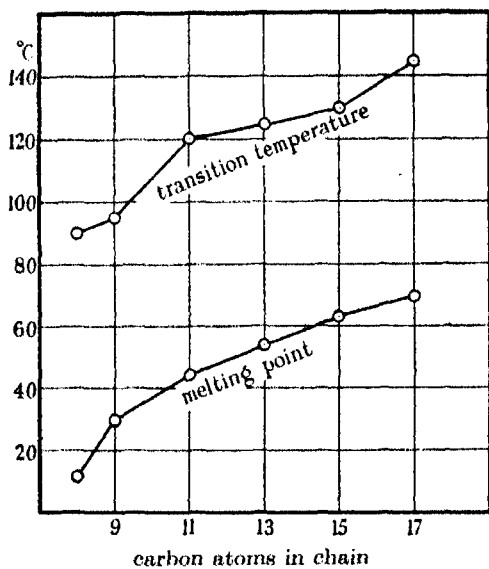
FIGURE 5. Fatty acids.
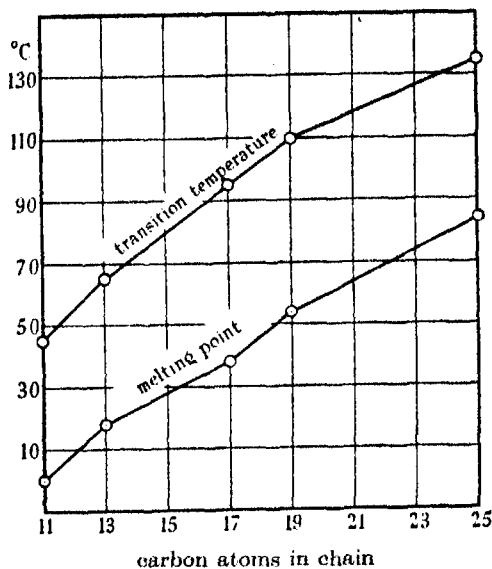
FIGURE 6. Methyl esters of fatty acids.
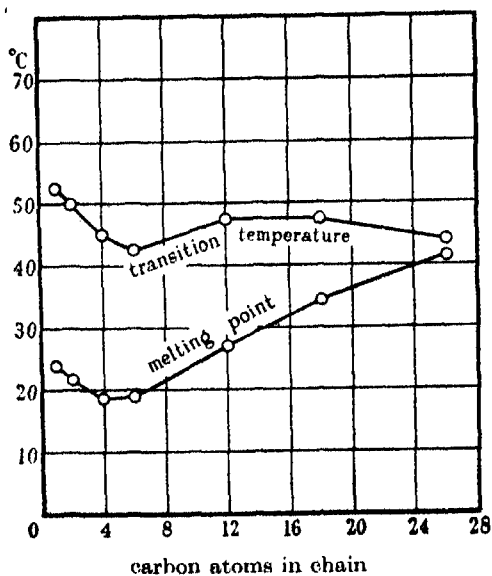
FIGURE 7. Stearic esters of normal alcohols.
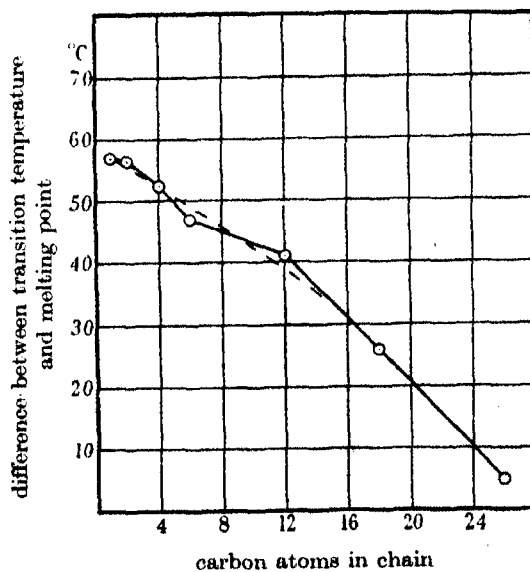
FIGURE 8. Stearic esters of normal alcohols.

Stereo-isomeric fatty acids

Some experiments were made in which oleic acid and its stereo-isomer elaidic acid were used as lubricants, and the following results were obtained :

	transition temp. °C	melting-point °C	difference °C
oleic acid (*cis*)	85	14	71
elaidic acid (*trans*)	120	45	75

Although these two acids differ in their configuration in space, the interval between the transition temperature and the melting-point is, within the limits of experimental error, the same in each case, and the same as that occurring with the saturated fatty acids.

Some experiments were also made with dicarboxylic fatty acids and aromatic acids, but these attacked the steel surface rapidly after melting.

Esters

Preparation and purification. The methyl esters, with the exception of methyl cerotate, were prepared by esterification with sulphuric acid and a large excess of methyl alcohol. Any free acid remaining was removed and the ester repeatedly recrystallized or fractionally distilled in vacuo. Methyl cerotate was prepared from dry silver cerotate and methyl iodide (cf. Whitby 1926).

The stearic esters of the normal alcohols were prepared from silver stearate and the iodide of the alcohol (cf. Whitby 1926). The difficulty in purifying long-chain compounds is well known (cf. Smith 1938), but the sample of stearic acid used in these preparations melted at 69° C and probably had a purity of at least 95 % (cf. Guy & Smith 1939). In the preparation of octadecyl stearate, octadecyl alcohol was first prepared by the reduction of ethyl stearate with sodium and alcohol as described by Smith (1931).

In the preparation of tristearin, stearyl chloride was first prepared by refluxing stearic acid with thionyl chloride on a water-bath and then fractionating the product in vacuo. Dehydrated glycerol dissolved in pyridine was added to a slight excess of stearyl chloride dissolved in petroleum ether. The mixture was allowed to stand for an hour and then refluxed for several hours on the water-bath. Enough petroleum ether was added to dissolve the tristearin and the mixture repeatedly extracted with dilute hydrochloric acid to remove the pyridine, and 10 % potassium carbonate solution containing some alcohol to remove any free stearic acid. The tristearin which was obtained on evaporating the petroleum ether was recrystallized from this solvent. It melted sharply at 71° C. Trimyristin, glycol distearate and glycol dimyristate were prepared in a similar manner.

The methyl esters of straight-chain fatty acids. The results for this series of esters are given in figure 6, from which it will be seen that the difference between the transition temperature and the melting-point is approximately 50° C for each member of the series.

The stearic esters of the normal alcohols. In this series of esters the effect of increasing the size of the 'alcohol' group from the methyl radical to the ceryl chain of 26 carbon atoms has been investigated. The results are given in figure 7. Both the melting-point and the transition temperature pass through a minimum in the region of C_4 to C_6 as the length of the 'alcohol' chain increases. Figure 8 shows, however, that the difference between the transition temperature and the melting-point steadily decreases with increase in the length of the alcohol chain and tends to become linear above about 16 carbon atoms.

Myristic and stearic esters of glycol and glycerol. The glycerol esters were investigated in view of the fact that most fatty oils used commercially as lubricants consist of the glycerol esters of long-chain fatty acids. The glycol esters may be regarded as intermediate in structure between the methyl and glycerol esters. The results obtained are given in Table 1, which includes those of the methyl esters for comparison.

TABLE 1

ester	transition temp. °C	melting-point °C	difference °C
methyl myristate	65	18	47
glycol dimyristate	85	61	24
glycerol trimyristate	80	56	24
methyl stearate	95	38	57
glycol distearate	115	76	39
glycerol tristearate	105	71	34

The methyl esters of dibasic fatty acids. Dimethyl succinate and dimethyl sebacate were examined in this series.

The results are as follows:

ester	transition temp. °C	melting-point °C	coefficient of friction	
			solid	liquid
dimethyl succinate	55	19·5	——	0·37
dimethyl sebacate	35	28	0·20	0·33

The friction with dimethyl sebacate as a lubricant rises rapidly as the surface is heated above the melting-point of this lubricant. A photographic record showing this effect is reproduced in figure 9. No change in friction was observed at the melting-points of the monobasic acids and their esters (cf. figure 3). The increase in friction at the melting-point of dimethyl sebacate suggests that a change in the orientation of the molecules in the surface layer occurs at the melting-point. Below the melting-point the molecules are probably attached to the surface by one of the ester groups with the hydrocarbon chains pointing away from the surface and tightly packed together as in the crystal lattice. In this orientation the coefficient of friction would be of the same order as that found for long-chain monobasic acids and esters, which is the case. When the bulk melting-point is reached the molecules

have sufficient energy to break away from the configuration in which they exist in the crystal lattice, and it is possible that both ester groups are now attracted to the surface, causing the molecules to lie more or less flat on the surface (cf. Adam & Jessop 1926). The higher value of the coefficient of friction in the liquid state supports this view. If the molecules lie flat on the surface above the bulk melting-point, it might be supposed that the adsorbed film of the succinate would be maintained to a higher temperature than that of the sebacate in which the film would be weakened by the thermal energy of the longer hydrocarbon chain joining the ester groups. The transition temperature of the succinate would therefore be higher than that of the sebacate which is the result observed.
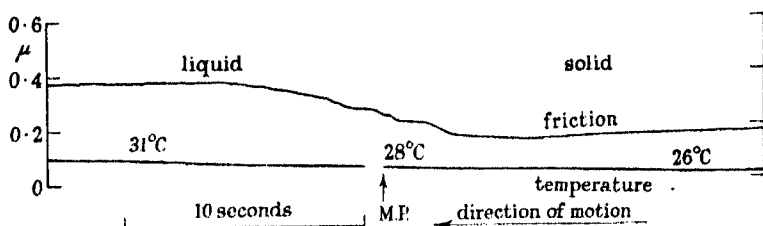
FIGURE 9. Dimethyl sebacate.

Discussion

Using a load of 4 kg. on the hemispherical contact and a sliding speed of 0·005 cm./sec., all the pure substances used as lubricants gave a transition from smooth sliding to stick-slips as the surface was heated. This transition occurred at the bulk melting-point of the hydrocarbons, ketones, alcohols and amides when the mild steel surfaces were lubricated with these substances, but at a considerably higher temperature than the bulk melting-point when acids and esters were used as lubricants.

In order to explain this phenomenon which, from the large number of substances investigated, appears to be quite general, the following hypothesis is advanced: Under the conditions of 'boundary lubrication' obtaining in the test, smooth sliding occurs when the surface layer of lubricant on the steel contact slides over a similar layer on the steel plate. Below the bulk melting-point the molecules are maintained oriented in the surface layer by the forces which maintain them in position in the crystal lattice even if, like the hydrocarbons, they contain no polar group which is attached to the surface. Smooth sliding therefore occurs below the bulk melting-point of all the substances examined. When the surfaces are heated up a temperature will be reached when the molecules in the surface layer become disordered and the layer 'melts'. When this occurs the metal surfaces are no longer protected by the oriented surface film and metal to metal contact can occur. The contact becomes stuck to the surface and travels with it until the limiting value of the static friction is reached. A slip then occurs under the action of the restoring force of the bifilar suspension. The contact then sticks to the surface again and

the process is repeated. Since hydrocarbons are non-polar and therefore not adsorbed, the surface layer of these compounds breaks down at the melting-point. Acids and esters, on the other hand, being strongly polar substances are adsorbed on the steel surface. The surface layer of acids and esters does not break down at the bulk melting-point, but is maintained to a much higher temperature. The transition temperature represents the temperature at which the adsorbed film breaks down or melts, and may therefore be regarded as a measure of the strength of this film. The difference between the transition temperature and the bulk melting-point which is approximately constant for the long-chain fatty acids and for their methyl esters, may be regarded as a measure of the strength with which the polar group is adsorbed, since the molecules have sufficient energy at the transition temperature to overcome both the lateral adhesion which holds them in position in the crystal lattice, and the force of adsorption of the polar group for the surface. Further experimental evidence in support of this hypothesis is afforded by the results obtained with built-up films described in the next part of this paper.

Evidence of the existence of an adsorbed film of lubricant at a metal surface and its persistence to elevated temperatures is afforded by the electron diffraction experiments of other workers. Andrew (1936) demonstrated the existence of adsorbed films at room temperature in the case of commercial oils of the type which give smooth sliding. Tanaka (1938, 1939) showed that orientation of palmitic and stearic acid on copper surfaces was maintained to 120–130° C. Beeck et al. (1940) found that the hydrocarbon tetratriacontane and the ketone stearone showed a high degree of orientation perpendicular to the surface below their melting-points, but the adsorbed molecules evaporated from the surface near the melting-point. An adsorbed layer of stearic acid, on the other hand, persisted at 110° C.

Since alcohols, ketones and amides, like hydrocarbons, give a transition to stick-slips at their melting-point, they cannot be strongly adsorbed on a steel surface. Esters and acids which are strongly adsorbed differ from alcohols, ketones and amides in containing the group —CO.O—. The fact that the methyl esters are almost as strongly adsorbed on a steel surface as the acids suggests that the adsorption is due to the interaction of the dipoles of the —CO.O— group with the atoms in the metal surface, since there is no question of ionization—which might explain the adsorption of the acids—playing any part in the adsorption of the esters. Furthermore, it was shown in the case of the acids that the transition is not due to their attacking the surface.

As the length of the hydrocarbon chain attached to stearic acid in the stearic esters of normal alcohols is increased, the difference between the transition temperature and the melting-point, and hence the strength with which the polar group is adsorbed, decreases as shown in figure 8. There are two factors which may account for this behaviour:

(1) The strength of adsorption may be weakened as the length of the 'alcohol' chain increases: the larger the 'alcohol' group the further the —CO.O— group may be kept from the surface.

(2) As the length of the 'alcohol' chain increases, the lateral adhesion between it and the 'acidic' chain may distort the molecule, and in the case of the long 'alcohol' chains, bend it into the form of a hairpin. This distortion which will increase with the length of the 'alcohol' chain will alter the configuration of the polar group and consequently change its resultant dipole moment.

A similar series of esters of palmitic acid has been studied as unimolecular films on aqueous surfaces by Adam (1929), and by Alexander & Schulman (1937), who have shown that if the 'alcohol' chain contains four or fewer carbon atoms it is forced under the surface and vertically opposes the 'acidic' chain in the condensed films. If the alcohol chain contains more than four carbon atoms, the lateral adhesion between it and the acidic chain is sufficient to bend the molecule up into the form of a hairpin. Alexander & Schulman find that in the case of cetyl palmitate the distortion of the polar group is sufficient to reduce the vertical component of the dipole moment to zero.

It is not improbable that the series of stearic esters are orientated in a similar manner on a steel surface, in which case a qualitative explanation is afforded of the results.

Further evidence in support of the view that the adsorption of acids and esters is due to the interaction of the dipoles of the —CO.O— group with the atoms in the metal surface is afforded by an experiment with octadecyl acetate as lubricant. This substance may be regarded as being derived from the methyl esters of the long-chain acids with the position of the hydrocarbon chain and the methyl radical interchanged. The position of the —CO.O— group in space in the adsorbed layer would therefore be different from what it is in the case of the methyl esters, resulting in a different value for the vertical components of the resultant dipole moment and hence in the strength of adsorption. The difference between the transition temperature and the melting-point of octadecyl acetate is 38° C compared with 51° C, the mean value for the methyl esters.

If the glycol and glycerol esters of long-chain acids are adsorbed on a steel surface in the same way as Adam (1922) has shown them to be oriented on a water surface —with the polar group on the surface and the hydrocarbon chains tightly packed and pointing away from it—they may be regarded as derived from two or three molecules respectively of the corresponding methyl esters linked through the carbon atoms of the methyl groups. The glycol esters have two polar groups in the molecule, the dipoles of which can interact with the metal atoms in the surface and cause the molecule to be adsorbed on the surface. These polar groups are the same as those in the methyl esters. Since there are two of these groups in the glycol esters it might be expected that the molecule of the glycol ester would be twice as strongly adsorbed as that of the methyl ester, but the presence of the additional polar group is offset by an additional hydrocarbon chain which can take up energy and contribute to the removal of the molecule from the surface at the transition temperature. If the chains are tightly packed distortion from the normal valency angles must occur. The resultant distortion at the polar groups would, as in the

case of the stearic esters, diminish the strength with which the molecules are adsorbed on the surface. A similar distortion would occur in the glycerol esters. The difference between the transition temperature and the melting-point would therefore be less for the glycerol and glycol esters than for the methyl esters, which is the result found experimentally.

II. Experiments with built-up films of lubricants

In order to test further the hypothesis that the transition from smooth sliding to stick-slips occurs when the adsorbed film becomes disoriented, the following experiments were made with lubricant films of known molecular thickness built up by the Langmuir-Blodgett'(1935) technique. Stainless steel surfaces were used in these experiments, since the mild steel previously employed was corroded by immersion in water. The films of the acids were spread on the surface of distilled water to avoid the deposition of the calcium salts which occurs when tap water is used.

The first unimolecular layer is deposited with the polar group adjacent to the surface, and thereafter the film is built up in double layers in which the polar groups are adjacent as in the crystal lattice. According to Stenhagen (1938) the long spacing of these built-up films as measured by X-rays always corresponds to a crystalline form of the substance, and Bikerman (1938) concludes that crystallization occurs in the film so that the surface becomes coated with a layer of microcrystals. The area of film deposited corresponds to the geometrical area of the surface on which it is deposited, although in the case of metals the actual area, due to pits and scratches, must be considerably greater than the geometrical area. It must therefore be supposed that the film is stretched over the surface like a soap film on wire gauze. In this connexion it is interesting to note that Bikerman (1939) finds that films may be deposited on wire gauze, the area of film deposited being equal to the geometrical area of the gauze.

Ceryl alcohol and myristone films

The mean values of a number of determinations of the transition temperature from smooth sliding to stick-slips, and the coefficient of friction below the transition temperature (in the region of smooth sliding) for ceryl alcohol and myristone films of varying thickness are given in table 2. When the surfaces were each coated with a unimolecular film of ceryl alcohol and heated up at a uniform rate, the friction remained steady until the bulk melting-point was reached, when small stick-slips occurred. These stick-slips persisted until the temperature had risen to about 100° C, when the friction became irregular, eventually rising to the high value of the clean metals. Evidently the film melted when the stick-slips set in, and remained liquid on the surface until a higher temperature was reached and the molecules had sufficient energy to evaporate from the surface. When five or more

molecular layers were present initially the stick-slips persisted to a higher temperature. The behaviour of the myristone films was essentially similar.

TABLE 2

lubricant	thickness of film in molecules	transition temp. °C	coefficient of friction
ceryl alcohol (m.p. 80° C)	1	79	0·11
	5	80	0·10
	15	79	0·09
	excess	82	0·10
myristone (m.p. 76° C)	1	75	0·17
	5	75	0·15
	excess	76	0·16

Methyl stearate, ethyl stearate, elaidic acid and stearic acid films

When the stainless steel surfaces were each coated with a unimolecular film of these substances and the friction observed as the surface was heated up from room temperature at a uniform rate, the friction remained steady until a temperature characteristic of the lubricant was reached. A sudden rise in friction then occurred which was recorded by the camera as a rise in the friction trace. A typical example of this first rise in friction in the case of elaidic acid is shown in figure 10a. The
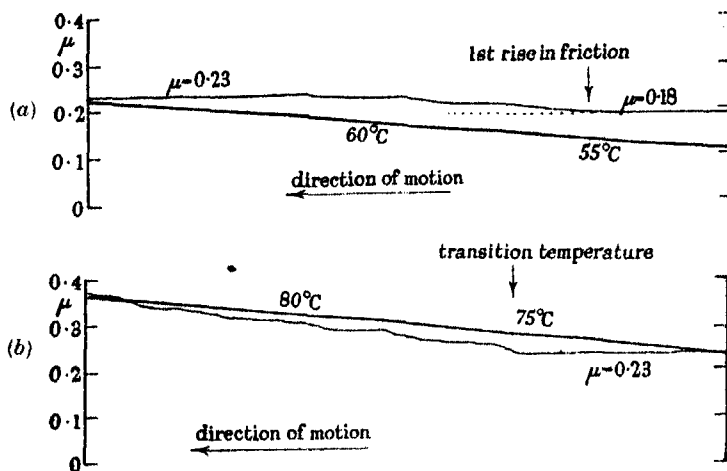
FIGURE 10. Unimolecular film of elaidic acid.

friction then remained steady at its new higher value until a temperature was reached at which the film broke down. When this occurred a rise in the friction, eventually to the high value, of the clean metals occurred. The photographic record of this is shown in figure 10b, which is the continuation of the record of figure 10a. The temperature at which this second rise in friction occurs is taken as the transition temperature. Since the adsorbed film is at a temperature considerably above

the bulk melting-point of the lubricant when it breaks down, the molecules do not persist in a disoriented condition but evaporate and the friction therefore changes from smooth sliding, not to stick-slips, but to the higher irregular value observed with unlubricated metals.

The mean results for unimolecular films of these lubricants are given in table 3. By analogy with the behaviour of films on aqueous surfaces it is probable that the first rise of friction is due to the surface films changing from the 'condensed' to the 'liquid-expanded' type. In the condensed type of film the molecules are closely packed and strongly attached to one another by the lateral adhesion

TABLE 3

lubricant (present as a unimolecular film on each surface)	coefficient of friction below 1st rise in friction	1st rise in friction trace °C	coefficient of friction above 1st rise in friction	transition temp. °C
methyl stearate	0·14	25	0·18	50
ethyl stearate	0·15	30	0·18	60
stearic acid	0·15	55	0·18	95
elaidic acid	0·17	55	0·22	75

between the hydrocarbon chains. In the liquid-expanded film the hydrocarbon chains are generally supposed to be free to rotate and vibrate but are firmly

anchored to the surface at the base by their polar end-group. In view of the flexibility of the hydrocarbon chains in the 'liquid-expanded' films the metal surfaces would be able to approach each other more closely than when covered with 'condensed' films, and this fact probably accounts for the higher coefficient of friction observed above the temperature of the first rise in friction.

In another series of experiments the stainless steel contact was coated with a unimolecular film of lubricant, and films of varying thickness were built up on the stainless steel plate. The latter was then heated up at a uniform rate and the friction with the contact observed.
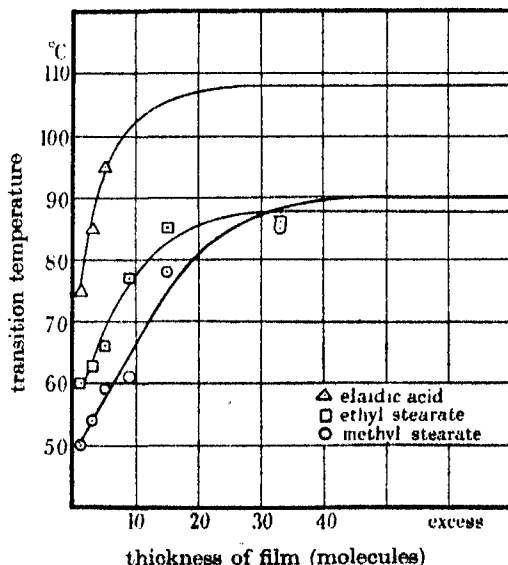
FIGURE 11

When five or more films were originally present on the surface a transition from smooth sliding to stick-slips occurred; with fewer films present initially the break-

down of the film was accompanied by the rise in friction mentioned above. The value of the transition temperature increased with the initial thickness of the film eventually reaching that recorded for an excess of lubricant on stainless steel surfaces. The results on methyl stearate, ethyl stearate, and elaidic acid are shown in figure 11.

The transition temperature for an excess of lubricant is higher than for a unimolecular film, since the adsorbed layer is in equilibrium with an excess of the molecules of which it is composed. Intermediate values of the transition temperature are obtained when the initial thickness of the film is up to about 30 mol. Owing to the short range over which adsorptive forces act the outer layers of lubricant will probably melt at the bulk melting-point, and will begin to evaporate from the surface as the temperature rises. A film which was initially 30 mol. thick will be considerably thinner by the time the transition temperature is reached. Experimental evidence of this is afforded by the fact that films of 33 molecular layers of methyl stearate, which are purple in white light, show no interference colour after being heated to 100° C. The film must therefore have been at least halved in thickness. The actual thickness of any film by the time the transition temperature is reached will depend on the rate of heating (which was constant in these experiments) and the volatility of the lubricant.

The effect of white oil on a unimolecular film of methyl stearate

In these experiments the stainless steel surfaces were each coated with a unimolecular film of methyl stearate and the lower surface set in motion, the friction being recorded by the camera. After the surface had been moving for about 30 sec.
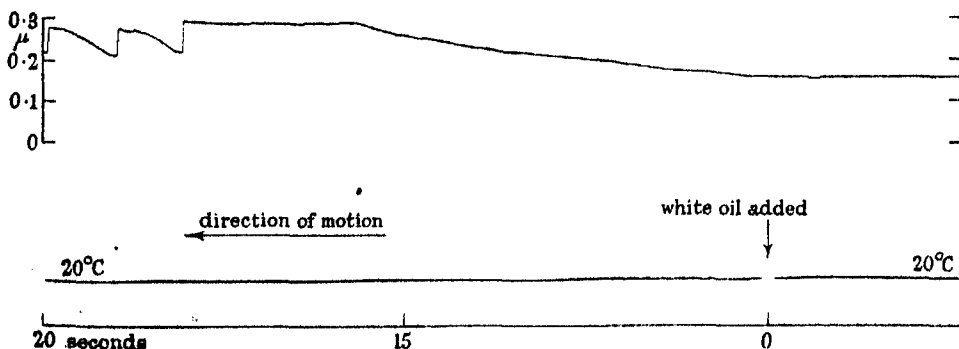
FIGURE 12. Unimolecular film of methyl stearate.

a drop of white oil was blown on to the surface round the contact from a pipette, the light beam recording the temperature being momentarily interrupted so as to record the moment of addition of the oil on the photographic film. The friction between the surfaces was recorded on the camera for about 3 min. after the addition of the oil. A typical record is reproduced in figure 12. It will be seen that

the friction between the surfaces coated with a unimolecular film of methyl stearate remained constant before the white oil was added. Immediately after the addition of white oil the coefficient of friction. rose during 15 sec., and then remained approximately constant for about 7 sec., when a slip occurred, and was followed by regular stick-slips.

This result shows that a unimolecular film of methyl stearate on stainless steel is dissolved from the surface by white oil in about 15 sec. at 20° C, since at the end of this period the behaviour is practically the same as if the surfaces are lubricated with white oil. If the stick-slips were due to sliding over the methyl stearate film in the presence of white oil, they should commence immediately the white oil is added. But the slope of the rise in friction which occurs after the white oil is added to the surface is much less than that of the 'stick', so that there is relative motion between the contact and the surface. Immediately the white oil is added to the surface, solution of the methyl stearate commences, and as the film is gradually removed the friction rises but sliding still occurs. When the film has passed into solution the surfaces are in effect lubricated with while oil and stick-slips characteristic of this condition occur.

Some experiments were also made in which the surfaces were each coated with a unimolecular film of elaidic acid, and white oil added. Since elaidic acid is more strongly adsorbed than methyl stearate, smooth sliding occurs at room temperature in the presence of white oil. The variation with temperature of the friction of a unimolecular film of elaidic acid covered with white oil showed a resemblance to the behaviour of a unimolecular film of elaidic acid in the absence of white oil. The first rise in friction occurred in the neighbourhood of 50° C and the film broke down and passed into solution at about 60° C. The resultant solution of elaidic acid in white oil was of insufficient concentration to build up an adsorbed film on cooling and the behaviour remained essentially that of white oil on stainless steel.

Solution of lubricants in white oil

A 1 % solution of methyl stearate in white oil continued to give stick-slips characteristic of white oil after being on the stainless steel surfaces for 18 hr. A solution of this concentration is therefore unable to build up a continuous adsorbed film in 18 hr. With a 10 % solution of methyl stearate in white oil smooth sliding occurred after a few hours showing that a continuous adsorbed film had been formed. When the surfaces were heated up the film broke down and passed into solution in the neighbourhood of 30° C, stick-slips characteristic of white oil occurring. On cooling to room temperature the stick-slips persisted and several hours elapsed before smooth sliding occurred, indicating that a continuous adsorbed film had been rebuilt on the surfaces. This shows that the transition from smooth sliding to stick-slips is due to the disorientation of the adsorbed film and its passage into the solution and not just to a loosening of the film as the temperature rose, in which case smooth sliding should recommence on cooling below the transition temperature.

Similar results were obtained with solutions of oleic acid in white oil. The more concentrated the solution the more rapidly the surfaces became coated with a continuous adsorbed film.

Discussion

When stainless steel surfaces are each coated with a unimolecular layer of ceryl alcohol or myristone a transition from smooth sliding to stick-slips occurs at the bulk melting-point of the lubricant. The same result is obtained when an excess of either of these lubricants is present, showing that the transition is caused by the disorientation or 'melting' of the surface layer of lubricant. The surface layers of ceryl alcohol and myristone are not strongly adsorbed and, like the non-polar hydrocarbons, break down at the bulk melting-point when the molecules have sufficient energy to break away from their positions in the crystal lattice.

When acids or esters are used as lubricants the surface film is maintained oriented by the adsorption forces to a temperature higher than the bulk melting-point. When in contact and in equilibrium with an excess of the molecules of which it is composed, a unimolecular film of acids or esters can be maintained as a continuous adsorbed film to a higher temperature than when it is isolated on the metal surface.

The experiments on the effect of white oil on unimolecular films of elaidic acid and methyl stearate afford further confirmation of the hypothesis that the transition from smooth sliding to stick-slips, is due to the disorientation of the adsorbed layer.

The experiments on built-up films show that the coefficient of friction under high loads is the same whether the surfaces are each coated with a unimolecular film or with a film of lubricant many molecules thick. This result affords direct experimental evidence in support of the view that under conditions of 'boundary lubrication' prevailing when the pressure between slowly moving surfaces is very high, excess of lubricant is squeezed out and lubrication is effected by a unimolecular layer adsorbed on each surface. The variation with load of the coefficient of friction between steel surfaces lubricated with oleic acid shows that orientation extends beyond the primary adsorbed layer in the case of this lubricant. The outer layers are not so strongly held as the primary layer and are squeezed out under high pressures. In the case of the esters, alcohols and ketones no appreciable orientation appears to extend beyond the primary layer, but some preliminary experiments on elaidic and stearic acid suggest that orientation may extend to several molecules with these lubricants. This accounts for the low values of the coefficient of friction, between surfaces lubricated with oleic acid, reported by previous workers and explains the good lubricating properties of this substance. Since the length of the molecule of oleic acid is approximately the same as that of methyl stearate, ethyl stearate, myristone, and elaidic acid, the coefficient of friction under true boundary conditions should be approximately the same as that obtained with built-up unimolecular films of the latter lubricants. Owing to orientation extending beyond the primary layer this value is attained in the case of oleic acid only under high loads.

The experiments described in this paper show that much information concerning adsorbed films on metal surfaces may be obtained by a study of the frictional behaviour of the surfaces at low speeds and under high loads.

The author wishes to express his thanks to the Shell Refining and Marketing Company Limited for permission to publish this work, which was carried out in their laboratories.

REFERENCES

Adam 1922 *Proc. Roy. Soc.* A, **101**, 452.
Adam 1929 *Proc. Roy. Soc.* A, **126**, 366.
Adam & Jessop 1926 *Proc. Roy. Soc.* A, **112**, 376.
Alexander & Schulman 1937 *Proc. Roy. Soc.* A, **161**, 115.
Andrew 1936 *Trans. Faraday Soc.* **32**, 607.
Beeck, Givens & Smith 1940 *Proc. Roy. Soc.* A, **177**, 90.
Bikerman 1938 *Trans. Faraday Soc.* **34**, 800.
Bikerman 1939 *Proc. Roy. Soc.* A, **170**, 130.
Blodgett 1935 *J. Amer. Chem. Soc.* **57**, 1007.
Bowden, Leben & Tabor 1939 *Trans. Faraday Soc.* **35**, 900.
Bowden & Leben 1939 *Proc. Roy. Soc.* A, **169**, 371.
Bowden & Leben 1940 *Phil. Trans.* **239**, 1.
Bowden & Tabor 1939 *Proc. Roy. Soc.* A, **169**, 391.
Bragg, W. H. 1925 *Not. Proc. Roy. Instn.* **24**, 481.
Devaux 1924 *J. Phys. Radium,* **5**, 84 *S.*
Finch & Zahoorbux 1937 *Instn Mech. Engrs, Disc. Lubrication,* **2**, group 4, p. 295.
Fogg 1940 *Proc. Phys. Soc.* **52**, 239.
Guy & Smith 1939 *J. Chem. Soc.* p. 615.
Hardy, W. B. 1936 *Collected Works.* Cambridge.
Langmuir 1920 *Trans. Faraday Soc.* **15**, 68.
Lawrence 1938 *Trans. Faraday Soc.* **34**, 1.
Müller 1923 *J. Chem. Soc.* **123**, 2043.
Rayleigh 1918 *Phil. Mag.* **35**, 158.
Smith 1931 *J. Chem. Soc.* p. 802.
Smith 1938 *Ann. Rep. Chem. Soc.* **35**, 252.
Stenhagen 1938 *Trans. Faraday Soc.* **34**, 1328.
Tabor 1940 *Nature, Lond.,* **145**, 308.
Tabor 1941 *Nature, Lond.,* **147**, 609.
Tanaka 1938 *Mem. Coll. Sci. Kyoto,* A, **21**, 85.
Tanaka 1939 *Mem. Coll. Sci. Kyoto,* A, **21**, 378.
Trillat 1925 *C.R. Acad. Sci., Paris,* **180**, 280.
Trillat 1929 *J. Phys. Radium,* **10**, 32.
Trillat 1930 *Z. Phys.* **64**, 191.
Whitby 1926 *J. Chem. Soc.* p. 1458.

Non-central forces in the nuclear two-body problem

By W. Hepner and R. Peierls

(Communicated by P. A. M. Dirac, F.R.S.—Received 2 December 1941)

The nuclear two-body problem is discussed on the assumption that the range of the forces is small compared with the size of the deuteron, but without specific assumptions about the forces. Using the values for the binding energy of the deuteron and its electric quadripole moment from observation, and eliminating a possible ambiguity in the deuteron wave function by means of the observed magnetic moments, it is found that the formulae for the electric and magnetic photo-effect in the deuteron remain the same as for central forces, with only minor modifications. The same is found for the scattering of neutrons by protons at energies of a few MV or less.

1. Introduction

The nuclear two-body problem has been discussed very extensively on the assumption of central forces between the particles.† It was found that, provided the range of the forces is smaller than the radius of the deuteron (so that the two particles in the deuteron will be within range of each other's attraction only for a small fraction of time) it is possible to describe all observed phenomena in terms of a few parameters quite independently of the detailed nature of the force. In particular, it was found that the scattering of neutrons by hydrogen, the capture of slow neutrons and the photo-disintegration of the deuteron depend in first approximation only on the binding energy of the deuteron, and on the energy of the (virtual) level belonging to the singlet state.

Since then the situation has been vitally changed by the discovery of the electric quadripole moment of the deuteron (Kellogg *et al.* 1939 *a,b*), which proves that the forces are not central, i.e. that the potential energy of a pair of particles depends not only on their distance, but also on their relative direction in relation to the spin axis.

The modifications introduced by this development affect mainly the triplet state, since in the singlet state all directions in space are equivalent, and hence the force must be central. It is of interest to see how many of the results of the simpler theory can be maintained if the non-central character of the forces is taken into account.‡ In the present paper we propose to investigate this question, and we shall begin with discussing the limiting case of a very short-range force. Actually it is likely that the range is comparable with the radius of the deuteron, and we shall also discuss the modifications in our results caused by this. The conclusion arrived at is that the results that have been obtained with central forces apply generally with only minor modifications.

† For a list of references on this subject, cf. Bethe (1936), in particular §§ 11–19.
‡ In a recent paper, Rarita & Schwinger (1941) have treated the same problems for a specific interaction potential. For this special case they come to the same conclusions which we arrive at below.

The principle of the simplifications applicable for short-range forces (Wigner 1933; Bethe & Peierls 1935) is that the wave equation at close approach is practically independent of the actual energy of the system, since in that region the potential energy is very much larger than the net energy. Hence it may be assumed that the wave function for this inner region is independent of energy, and instead of solving the wave equation containing the unknown forces, we may instead impose on the wave function a boundary condition at the point where the forces begin to matter. This boundary condition is approximately independent of energy and therefore always nearly the same as in the case of the deuteron; hence information on it can be obtained from the known properties of the deuteron. In the case of non-central forces the wave equation is a set of two simultaneous equations for two functions, and hence we obtain a set of boundary conditions rather than a single condition. Our first task is therefore to obtain the form of these conditions.

2. Equations and boundary conditions

In the non-relativistic equation of the two-body problem we imagine the centre of gravity split off; the remainder is then a function of the relative co-ordinates of the two particles and two spin variables. After introducing polar co-ordinates we can introduce the following variables: the distance between the particles, r; the quantum number of the resultant angular momentum, J; the component of the latter in the z direction, M; the resultant spin quantum number, s; the orbital angular momentum, l. Of these, J and M must commute with the energy, if the forces depend only on the relative orientation of $\mathbf{s}$ and $\mathbf{r}$, and we are interested mainly in the cases $J = 0$ and $J = 1$. s will also commute with the energy if the latter depends symmetrically on the two spins, and we assume this to be the case. (This rules out, for example, the magnetic interaction of the two spins with the orbital moment of the proton, which is likely to be small.) l is not a 'proper' quantum number if the forces are non-central. Permitted values are

$$l = J, \ J+1, \quad \text{if } s = 1, \qquad l = J, \quad \text{if } s = 0,$$

so that for low J the functions we are interested in are

	J	s	l		J	s	l
$^1S_0$	0	0	0	$^3S_1$	1	1	0
$^3P_0$	0	1	1	$^3P_1$	1	1	1
$^1P_1$	1	0	1	$^3D_1$	1	1	2

Of these states $^1S_0$, $^3P_0$, $^1P_1$, $^3P_1$ belong to combinations of 'exact' quantum numbers exclusive to them, since states with even l do not combine with states of odd l. But the $^3S_1$ and $^3D_1$ states may combine. We denote r times the radial wave functions of these states by f and g, respectively, choosing the factors in their definition in such a way that the normalization condition is given by

$$\int f^2 dr + \int g^2 dr = 1. \tag{1}$$

Without interaction the equations are then

$$\frac{d^2f}{dr^2} + \epsilon f = 0, \quad \frac{d^2g}{dr^2} - \frac{6}{r^2}g + \epsilon g = 0, \tag{2}$$

where $\epsilon \hbar^2/M = E$ is the energy of relative motion, if M is the mass of proton *or* neutron. We assume that equations (2) hold provided r is greater than the range a of the forces. For $r < a$ interaction terms will appear including, in particular, coupling terms, which introduce g in the first, and f in the second equation. Such terms have, for example, been discussed by Bethe (1939 *a*, *b*) and Rarita & Schwinger (1941). The wave equations (2) for the radial functions must be supplemented by the condition that both f and g vanish at $r = 0$. This fixes two of the four arbitrary constants involved in the solution of two simultaneous differential equations of second order, leaving still two parameters available. In other words, if we dispose of the values of f and g at the end of the range, $r = a$, we can still satisfy the equations, but the derivatives df/dr and dg/dr will then be determined. They must be linear functions of $f(a)$ and $g(a)$, and hence there must hold two relations of the form

$$\frac{df}{dr} = Af + Bg, \quad \frac{dg}{dr} = Cf + Dg \qquad (r = a). \tag{3}$$

The coefficients A, B, C, D are independent of what particular solution of the wave equation is chosen, provided it satisfies the boundary condition at $r = 0$.

Moreover, we may assume them to be independent of energy, provided the energy does not change by an amount comparable with the potential energy for $r < a$. Since this has to be of the order of 30 MV or more, in order that a bound state for the deuteron should exist, we may assume the coefficients of (3) to be the same for the normal state of the deuteron ($E = -2\cdot2$ MV) as for small positive energy. The variation of the coefficients with energy will be discussed in § 7.

The continuity equation yields a relation between the coefficients. Since (3) still leaves two parameters in the wave function free, we may make the wave function complex, and then the total flux passing through a sphere of any radius must vanish. This requires that

$$f^*\frac{df}{dr} - f\frac{df^*}{dr} + g^*\frac{dg}{dr} - g\frac{dg^*}{dr} = 0 \tag{4}$$

must hold identically for every solution compatible with (3). This is the case only if A and D are real and

$$B = C^*. \tag{5}$$

We shall, moreover, assume that B is real. This holds if the interaction force is given by real terms in the equations, as is the case for all interactions that have so far been discussed. A complex B would not affect our discussion very greatly.

The relation between the functions f and g and the actual wave function is provided by the formula

$$\Psi = \frac{1}{r}\left\{c_{10M}\, Y_0^0(\theta,\phi)\,\chi_0(\sigma)f(r) + \sum_m c_{12Mm}\, Y_2^m(\theta,\phi)\,\chi_{M-m}(\sigma)g(r)\right\}, \tag{6}$$

where Y_l^m are as usual normalized spherical harmonics, χ_{m_s} is the normalized spin function for the triplet state with z-component m_s of the total spin, and c_{JlMm} are the transformation coefficients from the quantum numbers $slMm$ to $slJM$. They satisfy amongst others the relation

$$\sum_M c_{10M0} c_{12M0}^* = 0, \tag{7}$$

which can be proved, for example, by noting that if the square of the function (6) is summed over M, the result represents the mean density for all orientations of the deuteron and must thus have spherical symmetry.

3. The normal state of the deuteron

The normal state of the deuteron, of energy $-E_0$, must belong to wave functions satisfying (2) and remaining bounded at infinite distance. The solutions satisfying this condition are

$$f = ce^{-\alpha r}, \quad g = de^{-\alpha r}\left(\frac{1}{(\alpha r)^2} + \frac{1}{\alpha r} + \frac{1}{3}\right), \tag{8}$$

where c and d are constants and α is defined by

$$\alpha^2 = \epsilon_0 = \frac{M\,|\,E_0\,|}{\hbar^2}. \tag{9}$$

Inserting these functions in (3), and eliminating c and d,

$$\frac{A+\alpha}{B} = \frac{B}{D + \frac{2}{a}\phi}, \tag{10}$$

where ϕ stands for

$$\phi = 1 + \frac{\frac{1}{3}(\alpha a)^2\,(1+\alpha a)}{3 + 3\alpha a + (\alpha a)^2}. \tag{11}$$

For very short-range forces, $\alpha a \ll 1$, ϕ is practically 1.

From (10)

$$\alpha = \frac{|\,B\,|^2}{D + \frac{2}{a}\phi} - A. \tag{12}$$

The other known (virtual) state of the deuteron belongs to a $^1S_0$ wave function which is not coupled to any other, and in which the force must be of central symmetry. Hence the discussion of this state is not modified.

All other states, made up entirely of wave functions with $l > 0$, are likely to belong to much higher energies.

A relation containing c and d is provided by the normalization condition (1). The integral in (1) is to be extended over all space, and since the expressions (8) apply only for $r > a$, the contribution of the inner part $r < a$ is unknown. If this

contribution is negligible, as in the case of very short-range forces, we can evaluate (1) by means of (8) and obtain

$$\frac{c^2 e^{-2\alpha a}}{2\alpha}(1+bq^2) = 1, \tag{13}$$

where we have put for brevity

$$b = \frac{1}{9} + \frac{2}{3}\frac{(1+\alpha a)^2}{(\alpha a)^3}, \tag{14}$$

$$q = d/c. \tag{15}$$

The ratio q depends on the extent to which the $^3D_1$ state is contained in the wave function of the ground state. This contribution can be estimated by means of the quadrupole moment with which we shall deal in the next section.

In the case of central forces it could be argued that the contribution from the inner part must be small, since the range of integration is small and there was no reason to expect the wave function to be large near the origin.

In our case this is not so certain. For one type of non-central force, for example, which can be obtained from the meson theory of nuclear forces, and which was discussed by Bethe (1939 a, b), both f and g would have a singularity at the origin unless the forces were 'cut off' at a certain limiting radius. If this radius is assumed to be much smaller than a, the wave function would be appreciably larger at low radii than for $r > a$.

A generous estimate of the inside contribution can be obtained by replacing the functions f and g by their maximum values for $0 < r < a$. This gives the contribution

$$(f^2_{\text{max.}} + g^2_{\text{max.}})\, a.$$

Unless the forces are of the singular type with a very small cut-off radius, the maximum values are likely to be of the same order as the values at $r = a$, and the above upper limit would then become

$$c^2 e^{-2\alpha a}\left\{1 + q^2\left(\frac{1}{(\alpha a)^2} + \frac{1}{\alpha a} + \frac{1}{3}\right)^2\right\} a.$$

For $\alpha a = \frac{1}{2}$ this is comparable to the outside part of the integral. Hence (unless extreme assumptions are made about the forces) the inside part is at the most equal to the outer one and likely to be smaller.

4. The quadripole moment of the deuteron

Nordsieck (1940) derived from Rabi's experiments the value of

$$Q = 2 \cdot 74 \times 10^{-27}\, \text{cm.}^2$$

for the quadripole moment in units of the proton charge.

Using relative co-ordinates, the quadripole moment is defined as the average, for the state $M = 1$, of

$$\tfrac{1}{4}(3z^2 - r^2),$$

and by integration over the angles and summation over spin this can be reduced to

$$Q = \frac{1}{10} \int r^2 (fg\sqrt{2} - \tfrac{1}{2}g^2)\, dr. \tag{16}$$

Here the question of the contribution from the inner region arises again and we assume, as before, that this will be negligible. Because of the factor r^2 in (16) this is a less stringent requirement than for the normalization integral.

Inserting from (8) in (16) and using (13)

$$Q = \frac{1}{10} \int_a^\infty r^2 (fg\sqrt{2} - \tfrac{1}{2}g^2)\, dr = \frac{1}{\alpha^2} \frac{vq - uq^2}{1 + bq^2}, \tag{17}$$

where we have put for brevity

$$v = \frac{\sqrt{2}}{30}[5 + 4\alpha a + (\alpha a)^2], \quad u = \frac{1}{30}\left[\frac{3}{\alpha a} + \tfrac{37}{12} + \tfrac{7}{8}\alpha a + \tfrac{1}{8}(\alpha a)^2\right].$$

The quadratic (17) has the two solutions

$$q_{1,2} = \frac{1}{2}\frac{v}{bQ\alpha^2 + u} \pm \frac{1}{2}\sqrt{\left(\frac{v^2}{(bQ\alpha^2 + u)^2} - \frac{4Q\alpha^2}{bQ\alpha^2 + u}\right)}.$$

Using the abbreviation
$$\xi = Q\alpha^2,$$

this reads
$$q = \frac{1}{2}\frac{v}{b\xi + u} \pm \frac{1}{2}\sqrt{\left(\frac{v^2}{(b\xi + u)^2} - \frac{4\xi}{b\xi + u}\right)}. \tag{18}$$

This shows first of all that no solution is possible unless†

$$4\xi(b\xi + u) < v^2. \tag{19}$$

The right-hand side of (19) decreases with decreasing a, whereas the left-hand side increases. With the empirical value of Q ($\xi = 0.0144$) the inequality holds down to $a = 0.25$. Hence it follows that in actual fact the range of the forces cannot be shorter than 1.1×10^{-13} cm. This limit is rather lower than the value commonly adopted, but it is of interest since it follows from arguments concerning only the two-body problem, without using the 'charge-independence' hypothesis, and is thus independent of assumptions about the additivity of forces and other complications of the many-body problem.

This conclusion is not seriously affected by our neglecting the contribution from $r < a$, since this would be small for $\alpha a = 0.25$. Moreover, the contribution from the

† If q were complex, the last factor in (17) would have to be replaced by

$$\{\tfrac{1}{2}v(q + q^*) - u|q|^2\}/(1 + b|q|^2).$$

The imaginary part of q would thus tend to reduce Q and the inequality (19) would hold

inner part to the denominator of (17) will partly cancel, if not outweigh, that to the numerator.

The most popular value of αa at present is about 0·5, and with this the two roots (18) become

$$q_1 = 0\cdot044, \quad q_2 = 0\cdot66. \tag{20}$$

Of these two values, the lower one seems to be more plausible, since q determines the ratio of g to f at large distances, and, with the larger q, g would be much larger than f near $r = a$. It is difficult to imagine forces of a reasonable type that would produce this result. Arguments based on the magnetic moment will, in fact, allow us to exclude the value q_2 (cf. § 5).

In deriving (17) we had assumed that the inside contribution was negligible. We can now estimate this contribution roughly by the same method as for the normalization. Of the two terms in (16) the internal contribution of the second is smaller than that of the first if q is a small number; and hence it is an overestimate to omit the second term and replace f and g by their maximum values. If we may again assume these to be of the same order as the values at $r = a$, the upper limit for the inside contribution to the integral (16) becomes

$$Q_{\text{inside}} < \frac{\sqrt{2}}{30} a^3 f(a)\, g(a).$$

Inserting for $f(a)$, $g(a)$ from (8),

$$Q_{\text{inside}} < \frac{\sqrt{2}}{30} a^3 c^2 q e^{-2\alpha a} \left(\frac{1}{(\alpha a)^2} + \frac{1}{\alpha a} + \frac{1}{3} \right).$$

With $\alpha a = \tfrac{1}{2}$, $q = 0\cdot044$, and using (13) this would be

$$Q_{\text{inside}} < 4 \times 10^{-2} Q. \tag{21}$$

The limit (21) confirms that the effect of the inside part in the integral for Q is likely to have less effect than the increase of the normalization integral.

5. THE MAGNETIC MOMENT OF THE DEUTERON

We assume that the magnetic moment of the deuteron is the resultant of the spin moments of neutron and proton and the orbital moment of the proton. Then if we denote by x the contribution of the D wave function to the normal state of the deuteron, the usual rules of vector combination give for the deuteron

$$\mu_D = (\mu_P + \mu_N)(1 - \tfrac{3}{2}x) + \tfrac{3}{4}x\mu_0, \tag{22}$$

where μ_P and μ_N are the proton and neutron moments and μ_0 the nuclear magneton.

It is known (Kellogg *et al.* 1939 *a, b*) that, in units of μ_0, $\mu_D = 0\cdot855$, $\mu_P = 2\cdot785$, and hence

$$\mu_N = -\frac{1\cdot930 - 3\cdot428x}{1 - 1\cdot5x}. \tag{23}$$

The behaviour of μ_N as a function of x from (23) is shown in table 1.

Since the observed value of μ_N is about $-1\cdot935$ according to Bloch & Alvarez (1940), it follows that x must be a small quantity below $0\cdot1$, and it must certainly lie below $0\cdot56$, unless μ_N is in fact positive or very large.

TABLE 1

x	0	0·023	0·1	0·2	0·3	0·4	0·5	0·5630
μ_N	$-1\cdot930$	$-1\cdot910$	$-1\cdot867$	$-1\cdot721$	$-1\cdot639$	$-1\cdot397$	$-0\cdot864$	0
x	0·6	0·65	0·66	0·67	0·7	0·8	0·8406	1
μ_N	1·268	11·53	33·25	$-459\cdot3$	$-9\cdot392$	$-4\cdot062$	$-3\cdot648$	$-2\cdot996$

In terms of our variables (cf. (13))

$$x = \frac{bq^2}{1+bq^2},\tag{24}$$

and for the higher value of q ($0\cdot66$) which we found compatible with the quadripole moment, we would obtain $x = 0\cdot84$. The contribution to the normalization integrals from close distances ($r < a$) will require a slight correction both to the value of q and to equation (24), but the corrections cannot be enough to reduce x to a value compatible with the above limits. We therefore adopt the lower value of q ($0\cdot044$) for all further calculations and have then

$$x = 0\cdot023.\tag{25}$$

6. The photo-effect of the deuteron

The disintegration of the deuteron by γ-rays is a transition from the ground state Ψ_0 into a state in the continuous spectrum with energy E. The transition is caused by the electric and magnetic dipole moment in the direction of polarization of the γ-ray. Quadripole moments are negligible as compared to the dipole effects for not too high energies. The cross-sections for these processes are given by

$$\sigma = 8\pi^3 \frac{\nu}{c}(M_{0E})^2,\tag{26}$$

where

$$h\nu = E + E_0 = \frac{\hbar^2}{M}(k^2 + \alpha^2)\tag{27}$$

is the energy of the γ-ray and M_{0E} is the matrix element of the moment in the transition from the ground state, with energy $-E_0$, to the state with energy E, taking the wave function of the latter state normalized per unit energy.

(i) Photo-electric effect

Assuming the γ-ray to be polarized in the z-direction the matrix element of the electric moment in relative co-ordinates is

$$M_{0E}^{\text{el.}} = \tfrac{1}{2}e\int \Psi_0^* z \Psi_E d\tau \quad (z = r\cos\theta).\tag{28}$$

Owing to Laporte's rule transitions from the ground state, which is even, into a state Ψ_E in the continuous spectrum are possible only if the latter is a P-state ($l = 1$) or an F-state ($l = 3$) or a mixture of both. For low energy E, as is the case in the experiments, the centrifugal forces are then already so large that the short-range interaction forces do not get any appreciable chance to influence the wave functions Ψ_E, which we may therefore assume to be wave functions of free particles. Also, the wave functions of the F-state will then be much smaller than that of the P-state, so that we can confine ourselves to transitions into the P-state of free particles. Since the electric moment commutes with the total spin, we need consider only the *triplet* P-state. The wave function of the latter, normalized per unit energy, is

$$\Psi_E = U_E^{(1)} Y_1^m(\theta,\phi)\,\chi = \left(\frac{M}{\pi\hbar^2 k}\right)^{\frac{1}{2}} \frac{1}{kr^2} Re\{e^{ikr}(-i-kr)\}\, Y_1^m(\theta,\phi)\,\chi, \qquad (29)$$

$Y_1^m(\theta,\phi)$ being the normalized spherical harmonics and χ the normalized triplet spin function. $Re\{\ldots\}$ stands for 'real part of'. The wave number k is defined by

$$\frac{\hbar^2}{M} k^2 = E. \qquad (30)$$

As before, we shall neglect the contribution from small radii ($r < a$) both in the matrix element and in the normalization of the ground state.

Integrating over the angles and summing over the spins, (28) becomes

$$M_{0E}^{\text{el}} = \tfrac{1}{2} e\left\{ c_{10Mm} \int_a^\infty f U_E^{(1)} r^2\, dr + \left(\frac{4-m^2}{5}\right)^{\frac{1}{2}} c_{12Mm} \int_a^\infty g U_E^{(1)} r^2\, dr \right\}, \qquad (31)$$

where c_{JlMm} is the transformation coefficient introduced in (6). Inserting for the wave functions from (8), (13), (15), (29), we have then

$$M_{0E}^{\text{el}} = -\frac{e}{2}\left(\frac{M}{3\pi\hbar^2 k^3}\right)^{\frac{1}{2}} \left(\frac{2\alpha}{1+bq^2}\right)^{\frac{1}{2}} e^{\alpha a}\left\{ c_{10Mm} Re \int_a^\infty e^{(ik-\alpha)r}(i+kr)\, dr \right.$$

$$\left. + q\left(\frac{4-m^2}{5}\right)^{\frac{1}{2}} c_{12Mm} Re \int_a^\infty e^{(ik-\alpha)r}(i+kr)\left(\frac{1}{(\alpha r)^2} + \frac{1}{\alpha r} + \frac{1}{3}\right) dr \right\}. \qquad (32)$$

With $ka \ll 1$, as assumed, and neglecting powers of ka and $(\alpha a)^2$ this becomes

$$M_{0E}^{\text{el}} = \frac{e}{2}\left(\frac{M}{3\pi\hbar^2}\right)^{\frac{1}{2}} \left(\frac{2\alpha}{1+bq^2}\right)^{\frac{1}{2}} (1+\alpha a) \frac{k^{\frac{1}{2}}}{(\alpha^2+k^2)^2}$$

$$\times \left\{ 2c_{10Mm} + \frac{5}{3}\left(1 + \frac{3}{5}\frac{k^2}{\alpha^2}\right) q\left(\frac{4-m^2}{5}\right)^{\frac{1}{2}} c_{12Mm} \right\}. \qquad (33)$$

Inserting in (26), summing over the components m of the final state and averaging over the orientation M of the initial state, we obtain, using (27),

$$\sigma_{\text{el.}} = \frac{8\pi}{3}\frac{e^2}{\hbar c}\frac{\hbar^2}{M}\frac{E_0^{\frac{1}{2}} E^{\frac{1}{2}}}{(E+E_0)^3}(1+\alpha a)^2 \frac{1}{1+bq^2}\left\{ 1 + \tfrac{5}{18}q^2\left(1 + \frac{3}{5}\frac{E}{E_0}\right)^2 \right\}. \qquad (34)$$

The term of first order in q vanishes because of the relation (9). With $\alpha a = 0\cdot5$, i.e. $b = 12\cdot1$, and $q = 0\cdot044$ (cf. (20)), this gives

$$\sigma_{\text{el.}} = 0\cdot98 \frac{8\pi e^2}{3} \frac{E_0^{\frac{1}{2}} E^{\frac{3}{2}}}{\hbar c\,(E+E_0)^3} (1+\alpha a)^2, \tag{35}$$

the contribution from the term in q^2 being negligible.

The corresponding calculation for central forces gives (Bethe 1936, §16, formula (79)† and Hall 1936)

$$\sigma_{\text{el.}} = \frac{8\pi e^2}{3} \frac{E_0^{\frac{1}{2}} E^{\frac{3}{2}}}{\hbar c\,(E+E_0)^3} (1+\alpha a)^2. \tag{35'}$$

We see thus that the effect of the non-central forces is practically negligible, producing merely a reduction of about 2 % in the magnitude of the effect owing to the decrease of the ^{3}S-function by this amount.

(ii) *Photo-magnetic effect*

This is a transition from the normal triplet state Ψ_0 of the deuteron into a singlet state Ψ_E of the continuous spectrum. The cross-section for the transition is again given by (26), M_{0E} denoting now the matrix element of the *magnetic* moment in the direction of polarization of the γ-ray

$$M_{0E}^{\text{magn.}} = \mu_0 \int \sum_{\sigma} \Psi_0 (\mu_P \sigma_P^z + \mu_N \sigma_N^z)\, \Psi_E \, d\tau, \tag{36}$$

where

$$\mu_0 = e\hbar/2Mc, \quad \mu_P = 2\cdot785, \quad \mu_N = -1\cdot935.$$

The orbital magnetic moment of the proton need not be included, since it gives no contribution to the transition from singlet to triplet state.

The spin operators in (36) act only on the spin co-ordinates of Ψ_E, i.e. on the spin function of the singlet state

$$\chi_s = \frac{1}{\sqrt{2}} \{\alpha(P)\beta(N) - \alpha(N)\beta(P)\}. \tag{37}$$

The non-central forces vanish in singlet states and therefore l is a good quantum number for the latter. The S-part of the deuteron ground state will therefore give rise to a transition into an S-state of the continuous spectrum, while the D-part will cause a transition into a D-state. In the latter the centrifugal force is so large that it is unaffected by the neutron-proton short-range force as long as the kinetic energy of the proton is not too high. We therefore take for it the wave function of a free

† Actually the formula quoted by Bethe for central forces was obtained by carrying all integrals down to $r = 0$, assuming the outside form of the wave function to hold even then. This gives no sense in our case because it would make g too strongly singular. If the integrals for central forces are extended only over radii greater than a, the formula quoted by Bethe is increased by a factor $(1+\alpha a)^2$. Hall includes the correct inside wave function for a specific potential function. His result is very nearly equal to (35′).

particle with angular momentum 2 and with wave number k corresponding to the energy E (cf. (30))

$$\Psi_E^{(2)} = U_E^{(2)} Y_2^m(\theta, \phi) \chi_s + \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \left(\frac{M}{2\hbar^2 k}\right)^{\frac{1}{2}} \frac{1}{r} Im\left[e^{ikr}\left(-1 - \frac{3i}{kr} + \frac{3}{(kr)^2}\right)\right] Y_2^m(\theta, \phi) \chi_s, \quad (38)$$

$Im(\ldots)$ standing for imaginary part of $(\ldots)$.

For the S-state we have to take into account the effect of the central force in the singlet state by writing

$$\Psi_E'^{(0)} = U_E^0 Y_0^0 \chi_s = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \left(\frac{M}{2\hbar^2 k}\right)^{\frac{1}{2}} \frac{1}{r} \sin(kr + \delta_0) Y_0^0 \chi_s, \quad (39)$$

where δ_0 is the phase shift as obtained from slow neutron scattering

$$\cot \delta_0 = +\beta/k, \quad (40)$$

β being the wave number corresponding to the energy E_0' of the singlet state of the deuteron, assuming this to be a virtual state.

We shall again neglect the contributions from close distances $(r < a)$ in the normalization of the ground state as well as in the matrix element (36). Integrating over the angles the latter becomes then

$$M_{0E}^{(0)\text{magn.}} = \mu_0 c_{10Mm} \sum_\sigma \chi_{M-m}(\mu_P \sigma_P^z + \mu_N \sigma_N^z) \chi_s \int_a^\infty f U_E^{(0)} r \, dr \quad (41)$$

for transitions into the singlet S-state,

and $\quad M_{0E}^{(2)\text{magn.}} = \mu_0 c_{12Mm} \sum_\sigma \chi_{M-m}(\mu_P \sigma_P^z + \mu_N \sigma_N^z) \chi_s \int_a^\infty g U_E^{(2)} r \, dr \quad (42)$

for transitions into the singlet D-state.

Since $\quad\quad \sigma_\nu^z \alpha(\nu) = \alpha(\nu), \quad \sigma_\nu^z \beta(\nu) = -\beta(\nu) \quad\quad (\nu = N, P),$

we obtain a non-vanishing matrix element only if

$$\chi_{M-m} = \frac{1}{\sqrt{2}}[\alpha(P)\beta(N) + \alpha(N)\beta(P)] = \chi_0,$$

i.e. if $\quad\quad\quad\quad\quad\quad\quad\quad M = m. \quad (43)$

Carrying out the summations over the spins and inserting the wave functions from (38), (39) and (8) we get respectively

$$M_{0E}^{(0)} = \mu_0(\mu_P - \mu_N)\left(\frac{M}{\pi\hbar^2 k}\right)^{\frac{1}{2}} \left(\frac{2\alpha}{1 + bq^2}\right)^{\frac{1}{2}} e^{\alpha a} c_{10MM} Im \int_a^\infty e^{(ik-\alpha)r + i\delta_0} \, dr, \quad (44)$$

$$M_{0E}^{(2)} = \mu_0(\mu_P - \mu_N)\left(\frac{M}{\pi\hbar^2 k}\right)^{\frac{1}{2}} \left(\frac{2\alpha}{1 + bq^2}\right)^{\frac{1}{2}} e^{\alpha a} q c_{12MM} Im \int_a^\infty e^{i(k-\alpha)r}$$

$$\times \left[-1 - \frac{3i}{kr} + \frac{3}{(kr)^2}\right]\left[\frac{1}{(\alpha r)^2} + \frac{1}{\alpha r} + \frac{1}{3}\right] dr \quad (45)$$

Integrating, and neglecting all powers of ka and of $(\alpha a)^2$,

$$M_{0E}^{(0)} = \mu_0(\mu_P - \mu_N)\left(\frac{M}{\pi\hbar^2}\right)^{\frac{1}{2}}\left(\frac{2}{1+bq^2}\right)^{\frac{1}{2}}\frac{\alpha^{\frac{1}{2}}k^{\frac{1}{2}}}{\alpha^2+k^2}c_{10MM}\frac{\alpha+\beta-\alpha a\left(\frac{k^2}{\alpha}-\beta\right)}{(\beta^2+k^2)^{\frac{1}{2}}}, \qquad (46)$$

$$M_{0E}^{(2)} = -\mu_0(\mu_P - \mu_N)\left(\frac{M}{\pi\hbar^2}\right)^{\frac{1}{2}}\left(\frac{2}{1+bq^2}\right)^{\frac{1}{2}}\frac{\alpha^{\frac{1}{2}}k^{\frac{1}{2}}}{\alpha^2+k^2}\tfrac{2}{3}qc_{12MM}(1+\alpha a)\left(1+\frac{1}{2}\frac{k^2}{\alpha^2}\right). \qquad (47)$$

Inserting in (26) and averaging over the orientation M of the initial state we obtain for the photo-magnetic cross-section, using (27),

$$\sigma = \frac{8\pi e^2}{3}\frac{\hbar^2}{\hbar c}\frac{E_0^{\frac{1}{2}}E^{\frac{1}{2}}}{4M^2c^2}\frac{1}{E+E_0}\frac{1}{1+bq^2}$$

$$\times\left\{\frac{\left[E_0^{\frac{1}{2}}+E_0'^{\frac{1}{2}}-\alpha a\left(\frac{E}{E_0^{\frac{1}{2}}}-E_0'^{\frac{1}{2}}\right)\right]^2}{E+E_0'}+\frac{4}{9}q^2(1+\alpha a)^2\left(1+\frac{1}{2}\frac{E}{E_0}\right)^2\right\}. \qquad (48)$$

We have here assumed that the singlet state of the deuteron is a virtual one (Schwinger & Teller 1934; Brickwedde, Dunning, Hoge & Manley 1938), otherwise the sign of the term in $E_0'^{\frac{1}{2}}$ should be reversed.

Of the two terms the first represents the transition into the S-state of the continuous spectrum, which is the same as for central forces corrected for finite range, except for a slight reduction due to the normalization factor. The second term represents the transition into the D-state and is new, but its contribution is too small to be observed.

7. Scattering

For positive energy the two solutions of the wave equation outside, normalized per unit energy, are

$$u_l = \left(\frac{\pi kr}{2}\right)^{\frac{1}{2}}J_{l+\frac{1}{2}}(kr), \qquad v_l = \left(\frac{\pi kr}{2}\right)^{\frac{1}{2}}J_{-l-\frac{1}{2}}(kr),$$

where

$$k^2 = ME/\hbar^2,$$

E being the energy of relative motion, i.e. one-half the energy of the incident neutron, J_p is the Bessel function of order p.

For $l = 0$ this gives f: $u_0 = \sin kr, \quad v_0 = \cos kr.$

For $l = 2$: g: $u_2 = -\sin kr - \dfrac{3\cos kr}{kr} + \dfrac{3\sin kr}{(kr)^2},$

$$v_2 = -\cos kr + \frac{3\sin kr}{kr} + \frac{3\cos kr}{(kr)^2}.$$

The solutions corresponding to the incident plane, plus emerging spherical, waves are

$$
\left.
\begin{aligned}
f &= I_0 \sin kr + S_0 e^{ikr}, \\
g &= I_2\!\left(-\sin kr - \frac{3\cos kr}{kr} + \frac{3\sin kr}{(kr)^2}\right) + S_2 e^{ikr}\!\left(-1 - \frac{3i}{kr} + \frac{3}{(kr)^2}\right)
\end{aligned}
\right\}.
\tag{49}
$$

For $kr \ll 1$ (denoting derivatives by accents):

$$
\left.
\begin{aligned}
f &= (I_0 + iS_0)\,kr + S_0 = S_0 + krI_0, \quad f' = k(I_0 + iS_0), \\
g &= \tfrac{1}{15}I_2(kr)^3 + 3S_2\frac{1}{(kr)^2}, \qquad\qquad g' = k\!\left\{\tfrac{1}{5}(kr)^2 I_2 - \frac{6}{(kr)^3} S_2\right\}.
\end{aligned}
\right\}
\tag{50}
$$

The boundary conditions are, according to (3) (assuming B to be real),

$$
f' = Af + Bg, \quad g' = Bf + Dg \qquad (r = a).
\tag{51}
$$

We assume that the values of the coefficients A, B, D in these equations are the same at a low positive energy E, as at the small negative energy $-E_0$. Then from (10)

$$
A + \alpha = \frac{B^2}{D + \dfrac{2}{\alpha}\phi}, \quad \text{with} \quad \phi = 1 + \frac{\tfrac{1}{2}(\alpha a)^2(1 + \alpha a)}{3 + 3\alpha a + (\alpha a)^2}.
$$

Inserting from (50) in (51)

$$
\left.
\begin{aligned}
\left(A - ik - \frac{B^2}{D + \dfrac{2}{a}}\right) S_0 &= \left(k - kaA + ka\frac{B^2}{D + \dfrac{2}{a}}\right) I_0 - \tfrac{1}{3}(ka)^2 k\frac{B}{D + \dfrac{2}{a}} I_2, \\
\left[(A - ik)\left(D + \frac{2}{a}\right) - B^2\right] S_2 &= \frac{(ka)^5}{45}\left[B^2 - \left(D - \frac{3}{a}\right)(A + ik)\right] I_2 - \tfrac{1}{3}(ka)^2 kBI_0.
\end{aligned}
\right\}
\tag{52}
$$

Neglecting terms of the order $(\alpha a)^2$ we have

$$
\frac{B_2}{D + \dfrac{2}{a}} = A + \alpha,
$$

hence

$$
\left.
\begin{aligned}
(-\alpha - ik) S_0 &= k(1 + \alpha a) I_0 - \tfrac{1}{3}(ka)^2 k\frac{B}{D + \dfrac{2}{a}} I_2, \\
\left(D + \frac{2}{a}\right)(-\alpha - ik) S_2 &= \frac{(ka)^5}{45}\left\{\left(D + \frac{2}{a}\right)\alpha + \frac{1}{a}(5A - 3ik)\right\} I_2 - \tfrac{1}{3}(ka)^2 kBI_0.
\end{aligned}
\right\}
\tag{53}
$$

From these equations it can be seen that the contribution of S_2 to the total scattering cross-section is negligible, since it is of the second order in ka. Similarly, in calculating S_0, we need not take into account the presence of the incident wave of $l = 2$ with amplitude I_2. With these approximations, the scattering cross-section becomes

$$
\sigma = 4\pi \left|\frac{S_0}{I_0}\right|^2 \frac{1}{k^2} = 4\pi \frac{(1 + \alpha a)^2}{\alpha^2 + k^2}.
\tag{54}
$$

This is the same formula as had been obtained previously for central forces, except for the expression in the numerator, which is connected with the finite range of the forces. Inasmuch as this expression differs from·unity, our approximation is not consistent, since the variation of the coefficients A, B, D with energy gives rise to terms of the same order.

We can now, however, deal with this energy dependence. For this purpose we note that the cross-section depends, apart from terms of the second order in ka, only on S_0, and that in calculating S_0 we may put $I_2 = 0$ in (49). Thus (49) reduces to

$$f = I_0 \sin kr + S_0 e^{ikr}, \quad g = - S_2 e^{ikr}\left(1 + \frac{3i}{kr} - \frac{3}{(kr)^2}\right). \tag{55}$$

Eliminating the arbitrary multiplicative constant which is involved in the definition of g, we have from continuity at $r = a$:

$$\frac{f'}{f} = \frac{kI_0 \cos kr + ikS_0 e^{ikr}}{I_0 \sin kr + S_0 e^{ikr}}. \tag{56}$$

We now determine the variation with energy of the left-hand side. One easily derives from the wave equations for the normal state of the deuteron and for a positive energy E, the result

$$\left\{ f\bar{f}\left(\frac{f'}{f} - \frac{\bar{f}'}{\bar{f}}\right) + g\bar{g}\left(\frac{g'}{g} - \frac{\bar{g}'}{\bar{g}}\right)\right\}_{r=a} = \frac{(E + E_0) M}{\hbar^2} \int_0^a (f\bar{f} + g\bar{g})\, dr. \tag{57}$$

Here f, g are, as before, the two wave functions of the deuteron, and $\bar{f}$, $\bar{g}$ those for positive energy. This equation holds if f, g is any solution of the wave equation at small distances, i.e. any linear combination of the two independent solutions. However, we are interested only in the case in which they are those solutions which actually describe the scattering, and this means that, in particular, $\bar{g}$ must, for $r \gtrless a$, satisfy the second equation (55). Hence

$$\frac{\bar{g}'}{\bar{g}} = \frac{i}{a} \frac{(ka)^3 + 3i(ka)^2 - 6(ka) - 6i}{(ka)^2 + 3ika - 3},$$

whereas from (8)

$$\frac{g'}{g} = -\frac{2}{a}\left\{1 + \frac{\frac{1}{2}(\alpha a)^2 (1 + \alpha a)}{3 + 3\alpha a + (\alpha a)^2}\right\}.$$

Hence, to first order in ka and αa

$$\frac{g'}{g} - \frac{\bar{g}'}{\bar{g}} = -\tfrac{1}{3}(k^2 + \alpha^2)\, a. \tag{58}$$

Inserting this in (57) and neglecting the second order of the difference between f, g and $\bar{f}$, $\bar{g}$,

$$\left(\frac{f'}{f}\right)_a - \left(\frac{\bar{f}'}{\bar{f}}\right)_a = \frac{(\alpha^2 + k^2)}{f(a)^2} \int_0^a (f^2 + g^2)\, dr + \tfrac{1}{3}(k^2 + \alpha^2)\frac{g(a)^2}{f(a)^2}a. \tag{59}$$

This expression is very similar to the one obtained for central forces (Bethe 1936, equation (56)) and the terms in g, which, in any case, would not affect the order of magnitude of the energy dependence, are in fact negligible. We can now work out a corrected cross-section from (56) which becomes

$$\sigma = \frac{4\pi}{k^2}\left|\frac{S_0}{I_0}\right|^2 = \frac{4\pi}{\alpha^2 + k^2}(1+\alpha a)^2(1-2\alpha a\gamma),\tag{60}$$

where

$$\gamma = \frac{1}{a[f(a)]^2}\int_0^a (f^2 + g^2)\,dr + \frac{1}{3}\frac{[1+\alpha a + \frac{1}{3}(\alpha a)^2]^2}{(\alpha a)^4}q^2.$$

As in the case of central forces,

$$\frac{1}{[f(a)]^2}\int_0^a (f^2 + g^2)\,dr$$

is of the order of magnitude of a, unless the functions f or g become singular near the origin. Hence the factor γ may be expected to be of order of magnitude unity, but it may be somewhat different from the value found by Bethe for a 'square-well' central force.

REFERENCES

Alvarez & Bloch 1940 *Phys. Rev.* **57**, 111.
Bethe & Peierls 1935 *Proc. Roy. Soc.* A, **148**. 146.
Bethe 1936 *Rev. Mod. Phys.* **8**, 80.
Bethe 1940*a* *Phys. Rev.* **57**, 260.
Bethe 1940*b* *Phys. Rev.* **57**, 390.
Brickwedde, Dunning, Hoge & Manley 1938 *Phys. Rev.* **54**, 266.
Hall 1936 *Phys. Rev.* **49**, 401.
Kellogg, Rabi, Ramsey, Jr. & Zacharias 1939*a* *Phys. Rev.* **55**, 318.
Kellogg, Rabi & Ramsey, Jr. 1939*b* *Phys. Rev.* **56**, 728.
Nordsieck 1940 *Phys. Rev.* **58**, 310.
Rarita & Schwinger 1941 *Phys. Rev.* **59**, 436.
Schwinger & Teller 1934 *Phys. Rev.* **52**, 286.
Wigner 1933 *Phys. Rev.* **43**, 252.

On the relation of Debye theory and the lattice theory of specific heats

By M. Blackman, *Imperial College, London*

In the Debye theory of the specific heat of solids the value of θ calculated from elastic constants, namely, θ_D (elastic), should be the same as that found from specific-heat data, θ_D (specific heat). The ratio of these θ values for crystal lattices is calculated here at a temperature $T \sim \theta_D$ on the basis of lattice theory; it is shown that the ratio is remarkably near unity for crystals of the NaCl type provided the crystals are not very anisotropic. In the case of other crystal types, the agreement is in general less good, the value of unity holding only in isolated cases.

Recent investigation into the vibrational spectra of solids (Blackman 1937; Fine 1939; Kellermann 1940) have all resulted in spectra which differ considerably from that assumed in the Debye theory of specific heats (Debye 1912). In view of these differences the agreement between Debye theory and experiment is surprisingly good, though there are a number of cases for which the theory does not hold (Eucken 1929).

One particular point of agreement is that the θ value calculated from elastic data at room temperature, θ_D (elastic), fits in well with the value of θ, θ_D (specific heat), obtained from the specific-heat curve at moderately high temperatures ($T \sim \theta_D$). Debye's calculation referred to an isotropic continuum, but a similar calculation can be made in the non-isotropic case; this extension of the original Debye theory is implied whenever a Debye formula is fitted to the specific-heat curve of a crystal.

The extent of the agreement is shown in table 1. It will be noted that the alkali halides give values of R [$R = \theta_D$ (el.)$/\theta_D$ (sp. ht.)] which are remarkably near unity.

TABLE 1

substance	θ from elastic data	from specific heat	$R = \dfrac{\theta_D \text{ (el.)}}{\theta_D \text{ (sp. ht.)}}$
NaCl	302	287	1·05
KCl	227	220	1·03
KBr	179	179	1·00
Cu	342	310	1·10
Ag	212	220	0·96
Au	158	186	0·85
W	384	310	1·21

A general agreement would follow of course from Debye theory, but is by no means an automatic consequence of the lattice theory of specific heats. The θ_D (el.) value is determined in the main by the mean velocity of long elastic waves, i.e. by

properties of the low-frequency vibrations of a crystal. The specific heat of a crystal at moderately high temperatures is determined by high-frequency vibrations. This specific heat is then translated into a value of θ_D (sp. ht.) at a given temperature by using a table of values of the Debye function. Thus the value of θ_D (sp. ht.) is actually determined by the high-frequency vibrations of the solid. It is therefore not at all clear that there should be any definite relation between the two values. Furthermore, an investigation of a simple cubic lattice with weak shear forces (Blackman 1935 b) indicated that the ratio R could be varied (by varying the force constant controlling the shear modulus). The theory was not worked out in detail, however, and it is the purpose of this paper to do so and also to investigate, in particular, the behaviour of the ratio for ionic crystals of the NaCl type.

The form which θ_D (el.) takes is discussed in § 1. The specific heat of a lattice at high temperatures is worked out in §2, and from the exact theory the value of a constant, θ_D (lattice), is obtained; the constant is the same as would be obtained from the specific heat of the lattice (at moderately high temperatures) when this is interpreted in terms of a Debye theory, i.e. the constant should be the same as θ_D (sp. ht.).

1. Once the elastic constants of a lattice are known, the θ_D (el.) value can be obtained by a variety of methods. These have been worked out particularly fully for cubic crystals. A very useful formula due to Born & v. Kármán (1914) holds for isotropic or nearly isotropic cases. This gives the correct mean velocity (c_m) of elastic waves in the form

$$\frac{3}{c_m^3} = \sigma^{\frac{3}{2}}\left\{\frac{2}{c_{44}^{\frac{3}{2}}} + \frac{1}{c_{11}^{\frac{3}{2}}} + \tfrac{3}{5}(c_{12} - c_{11} + 2c_{44})\left(\frac{1}{c_{44}^{\frac{5}{2}}} - \frac{1}{c_{11}^{\frac{5}{2}}}\right),\right. \tag{1}$$

where c_{11}, c_{12}, c_{44} are the three elastic constants of a cubic crystal, σ the density, and it is assumed that $|B| \ll 1$, where $B = (c_{12} - c_{11} + 2c_{44})/(c_{11} - c_{44})$. Once the mean velocity is found the θ values can be calculated from the expression

$$\theta_D\,(\text{el.}) = \frac{h}{k}c_m\left(\frac{3}{4\pi}\frac{N}{V}\right)^{\frac{1}{3}},$$

where V/N is the 'volume per particle', h Planck's constant and k Boltzmann's constant.

In the case where $c_{12}/c_{11} \ll 1$ and $c_{44}/c_{11} \ll 1$, an explicit formula for θ_D (el.) has been obtained (Blackman 1935b). The two regions in which these two formulae hold do not quite overlap, but, as will be seen below, all the information necessary for this investigation can be obtained without calculations in the intermediate region.

In the case of the simple cubic lattice, the relations for the elastic constants are the following (Born & v. Kármán 1912):

$$c_{11} = (\alpha + 4\gamma)/a, \quad c_{12} = c_{44} = 2\gamma/a,$$

where α is the force constant at a distance a (where a is the side of the unit cube), and γ is the force constant for particles at a distance $a\sqrt{2}$.

For ionic lattices of the NaCl type, the elastic constants have been worked out by Born (1920). I here follow the formulation given by Kellermann (1940) in which (besides the Coulomb terms) forces between nearest neighbours only are taken. The elastic constants have then the form

$$c_{11} = (A/2 - 2 \cdot 56)\, e^2/2r_0^4, \quad c_{12} = c_{44} = 0 \cdot 695 e^2/2r_0^4,$$

where e is the elementary unit of charge and r_0 the distance between nearest neighbours. The constant A is determined by the non-Coulomb forces (i.e. repulsive forces and the van der Waals forces) and is given by the relation

$$A = \frac{12 r_0^4}{\kappa e^4} + \tfrac{4}{3}\beta,$$

where $\beta\ (= 1 \cdot 7476)$ is Madelung's constant and κ the compressibility.

2. The calculation of θ_D at moderately high temperatures can be carried through by a method described by Thirring (1913). If ν is the frequency of a linear oscillator, the heat capacity can be expressed in the form

$$c_v = k\left\{ 1 + \sum_{n=1}^{\infty} (-1)^n B_n \frac{(2n-1)}{2n!} \left(\frac{h\nu}{kT}\right)^n \right\}$$

for values of $h\nu/kT < 2\pi$. In this expression B_n are the Bernoulli numbers, and the other constants have their usual meanings.

In order to obtain the heat capacity of a lattice, one has to sum this expression for all the frequencies of the lattice, i.e.

$$C_v = k\left(3N - \frac{B_1}{2} \frac{h^2}{k^2 T^2} \sum_{\nu} \nu^2 + \ldots \right)$$

$$= 3Nk\left(1 - \frac{1}{12} \frac{h^2}{k^2 T^2} \frac{1}{3N} \sum_{\nu} \nu^2 + \ldots \right), \tag{2}$$

where N is the total number of particles in the lattice. This expression can be compared with the corresponding Debye expression,

$$C_v = 3Nk\left(1 - \frac{1}{20} \frac{\theta_D^2}{T^2} + \ldots \right).$$

It is seen from the comparison that the specific heat can be described at moderately high temperatures by a parameter θ_D (lat.) given by

$$\theta_D^2(\text{lat.}) = \frac{5}{3} \frac{h^2}{k^2} \frac{1}{3N} \sum_{\nu} \nu^2. \tag{3}$$

This parameter θ_D (lat.) is then the same as the θ_D (sp. ht.) obtained by interpreting the heat capacity (2) in terms of the Debye theory.

Thirring showed that the values of $\Sigma\nu^2$, $\Sigma\nu^4$, etc., could be calculated once the equations of motion for a lattice are known. From these equations one obtains (Born & v. Kármán 1912) an equation for ν^2 in the form of a secular determinant. The value of $\Sigma\nu^2$, for example, can be obtained from the diagonal terms of the determinant.

In the case of the simple cubic lattice, the value of $\sum_\nu \nu^2$ can be obtained directly from Thirring's work, and

$$\theta_D^2 \,(\text{lat.}) = \frac{h^2}{k^2}\frac{5}{3}\frac{\alpha + 4\gamma}{2m\pi^2} = \frac{h^2}{k^2}\frac{5}{3}\frac{c_{11}a}{2m\pi^2},$$

where m is the mass of the particle in the unit (cubic) cell of side a.

The corresponding expression in the case of the ionic lattice can also be obtained in a compact form. In this case there are two particles in the unit cell; assuming central forces between particles, with a potential function $\psi(r)$, the equations of motion for the particles of the lattice have the form (Born 1923)

$$m_k \ddot{u}_{kx}^l - \sum_y \sum_{k'} \sum_{l'} (\psi_{kk'}^{l-l'})_{xy}\, u_{k'y}^{l'} = 0. \tag{4}$$

Here $k = 1, 2$ labels the two types of particles, l labels the cell, i.e. $l \equiv (l_1, l_2, l_3)$; u_{kx}^l is the displacement in the x direction, of the particle k in cell l, and

$$\psi(_{kk'}^{l-l'})_{xy} = \left(\frac{\partial^2 \psi_{kk'}(r)}{\partial x\, \partial y}\right)_{r = r_{kk'}^{l-l'}},$$

where $r_{kk'}^{l-l'}$ is the distance between the particles, k in cell l, and k' in cell l'. The case where $l = l'$, $k = k'$ has a separate definition (cf. Born 1923)

$$(\psi_{kk}^0)_{xy} = -\sum_{k'}\sideset{}{'}\sum_{l} (\psi_{kk'}^l)_{xy}, \tag{5}$$

Σ' indicating that the term $l = 0$, $k' = k$ is to be omitted.

Assuming periodic boundary conditions the plane-wave solutions of the equations of motion can be written in the form

$$u_{kx}^l = U_{kx}\, e^{-i\omega t}\, e^{i(l_1\phi_1 + l_2\phi_2 + l_3\phi_3)},$$

where ϕ_1, ϕ_2, ϕ_3 are the phases describing the normal vibration, and ω is the angular frequency; the equations of motion reduce to six equations of the type

$$\omega^2 m_k U_{kx} + \sum_{k'}\sum_y \begin{bmatrix} kk' \\ xy \end{bmatrix} U_{k'y} = 0, \tag{6}$$

where
$$\begin{bmatrix} kk' \\ xy \end{bmatrix} = \sum_l (\psi_{kk'}^l)_{xy}\, e^{i(l\phi)}. \tag{7}$$

These equations lead to a sixth-order determinant for ω^2, the diagonal terms of which have the form $m_k\omega^2 + \begin{bmatrix} kk \\ xx \end{bmatrix}$, and the non-diagonal terms the form $\begin{bmatrix} kk' \\ xy \end{bmatrix}$.

For a given set of values of ϕ_1, ϕ_2, ϕ_3 we obtain six frequencies, the sum of the squares of which is given by

$$\sum_{r=1}^{6} \nu_r^2 = -\frac{1}{4\pi^2 m_1}\left(\sum_{l,x}(\psi_{11}^l)_{xx}\, e^{i(l\phi)}\right) - \frac{1}{4\pi^2 m_2}\left(\sum_{l,x}(\psi_{22}^l)_{xx}\, e^{i(l\phi)}\right).$$

This expression has to be summed for all values of ϕ_1, ϕ_2, ϕ_3, giving in all $6N$ frequencies, if there are N particles of each type in the lattice. The summation over ϕ leads, however, to zero except in the case $l = 0$, i.e.

$$\sum_{\nu} \nu^2 = -\frac{N}{4\pi^2 m_1}\sum_{x}(\psi_{11}^0)_{xx} - \frac{N}{4\pi^2 m_2}\sum_{x}(\psi_{22}^0)_{xx}.$$

From the definition (5) it follows that

$$-(\psi_{11}^0)_{xx} = \sum_{l}(\psi_{12}^l)_{xx} + \sum_{l}{}'(\psi_{11}^l)_{xx}, \quad -(\psi_{22}^0)_{xx} = \sum_{l}(\psi_{21}^l)_{xx} + \sum_{l}{}'(\psi_{22}^l)_{xx}. \quad (8)$$

The Coulomb part of the potential leads to the same result for both expressions; if it is assumed that the non-Coulomb part of the potential is the same (i.e. irrespective of whether one is dealing with like or unlike particles), the two expressions become identical. This assumption appears to lead to reasonable results (cf. Kellermann 1940).

Under these conditions the sum of the squares of the frequencies becomes

$$\sum_{\nu} \nu^2 = -\frac{N}{4\pi^2}\left(\frac{1}{m_1}+\frac{1}{m_2}\right)\sum_{x}(\psi_{11}^0)_{xx}. \quad (9)$$

If we now group together the Coulomb terms in (9) we see (cf. (18)) that we have simply to sum the expression

$$\left(\frac{\partial^2 \psi}{\partial x^2}+\frac{\partial^2 \psi}{\partial y^2}+\frac{\partial^2 \psi}{\partial z^2}\right)_{\text{Coulomb}}$$

for all particles acting on one. This gives zero (since the Coulomb potential obeys the Laplace equation). We are hence left with the non-Coulomb part of the potential only, which we will call $v(r)$. Assuming that this falls off sufficiently rapidly, we are left with terms due to nearest neighbours only, i.e.

$$-\sum_{x}(\psi_{11}^0)_{xx} = 6\sum_{x}\left(\frac{\partial^2 v}{\partial x^2}\right)_{r=r_0} = 6\left(\frac{d^2 v}{dr^2}+\frac{2}{r}\frac{dv}{dr}\right)_{r=r_0},$$

r_0 being the distance between nearest neighbours. With the above assumptions as to the potential, the potential energy per cell, $\tfrac{1}{2}\Phi(r_0)$, can be written as

$$\tfrac{1}{2}\Phi(r_0) = -\beta e^2/r_0 + 6v(r_0),$$

where β is Madelung's constant. The equilibrium condition for the lattice leads to

$$\beta\frac{e^2}{r_0^2} + 6\left(\frac{dv}{dr}\right)_{r=r_0} = 0.$$

The second derivative of $v(r)$ can be expressed in terms of the compressibility κ of the lattice, which can be written as (Born 1923, p. 569)

$$\frac{1}{\kappa} = \frac{1}{18r_0}\left(\frac{d^2 \frac{1}{2}\Phi}{dr^2}\right)_{r=r_0} = \frac{1}{18r_0}\left\{-2\beta\frac{e^2}{r_0^3} + 6\left(\frac{d^2 v}{dr^2}\right)_{r=r_0}\right\}.$$

Hence
$$6\left(\frac{d^2 v}{dr^2} + \frac{2}{r}\frac{dv}{dr}\right)_{r=r_0} = \frac{18r_0}{\kappa}.$$

The relation (9) can now be written in the form

$$\sum_\nu \nu^2 = \frac{N}{4\pi^2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right)\frac{18r_0}{\kappa}.$$

In the case of an ionic lattice containing $2N$ particles, the heat capacity at high temperatures has the form

$$C_v = 6Nk\left(1 - \frac{1}{12}\frac{h^2}{k^2 T^2}\frac{1}{6N}\sum_\nu \nu^2 + \dots\right),$$

where $\sum_\nu \nu^2$ has the value given above. It follows that

$$\theta_D^2 \text{(lat.)} = \frac{5}{3}\frac{h^2}{k^2}\frac{1}{4\pi^2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right)\frac{3r_0}{\kappa}. \tag{10}$$

3. In the preceding section, formulae have been given for the specific heat of two types of cubic lattice at high temperatures, and the θ_D (lat.) values at high temperatures have been evaluated; the θ_D (el.) values were discussed in § 2. The calculation of the ratio of these two quantities is the main purpose of this paper, and the properties of this ratio will now be discussed.

It can be seen from (1) that the mean velocity c_m which occurs in the expression for θ (el.) can be written in the form

$$c_m = (c_{11}/\sigma)^{\frac{1}{2}} g(c_{12}/c_{11}, c_{44}/c_{11}), \tag{11}$$

where the function g is dimensionless. This expression holds in the general case also. (For the lattices with which we are concerned $c_{12} = c_{44}$, i.e. the function contains c_{12}/c_{11} only.)

In order to obtain the θ_D (el.) value from c_m, the procedure to be followed is that of Debye, namely, to integrate the density ($\rho(\nu)$) of normal vibrations (which depends on c_m) to a value ν_0 ($= k\theta_D$ (el.)$/h$), such that the correct number of normal vibrations is obtained. The density of normal vibrations has the form $\rho(\nu) = 12\pi V c_m^{-3}\nu^2$, where V is the volume of the crystal.

Considering the ionic lattices first, we take $V = N\varDelta$, where $\varDelta$ is the volume per cell ($\varDelta = 2r_0^3$); there are two particles per cell giving $2N$ particles in all, and $6N$ normal vibrations. Thus

$$6N = \int_0^{\nu_0} 12\pi N\varDelta c_m^{-3}\nu^2 d\nu = 8\pi N r_0^3 c_m^{-3}\nu_0^3,$$

or
$$\nu_0 = \frac{c_m}{r_0}\left(\frac{3}{4\pi}\right)^{\frac{1}{3}}, \quad \theta_D \text{(el.)} = \frac{h}{k}\frac{c_m}{r_0}\left(\frac{3}{4\pi}\right)^{\frac{1}{3}}.$$

Inserting the expression (11) for c_m and replacing the density σ by its value $(m_1 + m_2)/2r_0^3$ in this case, the expression for the square of θ_D can be written as

$$\theta_D^2 \,(\text{el.}) = \frac{h^2}{k^2}\left(\frac{3}{4\pi}\right)^{\frac{1}{3}}\frac{2c_{11}r_0}{(m_1 + m_2)}g^2(c_{12}/c_{11}). \tag{12}$$

Comparing (12) with (10) and noting that $3/\kappa = c_{11}(1 + 2c_{12}/c_{11})$ we see that c_{11} and r_0 occur in the same way in both expressions; hence the ratio R of the θ values depends on c_{12}/c_{11} and on m_2/m_1; the expression for R becomes

$$R = \frac{\theta_D(\text{el.})}{\theta_D(\text{lat.})} = \left(\frac{3}{4\pi}\right)^{\frac{1}{3}}\left(\frac{6}{5}\right)^{\frac{1}{2}}\pi\left(\frac{4m_1m_2}{(m_1 + m_2)^2}\right)^{\frac{1}{2}}\frac{g(c_{12}/c_{11})}{(1 + 2c_{12}/c_{11})^{\frac{1}{2}}}. \tag{13}$$

This function is plotted in curve (a) of figure 1 for equal masses. The rather unexpected feature is that R becomes practically constant (and equal to 1·06) between $c_{12}/c_{11} = 0·20$, and the extreme value calculated $c_{12}/c_{11} = 0·40$. This range includes practically all crystals of the NaCl type for which data are available. The effect
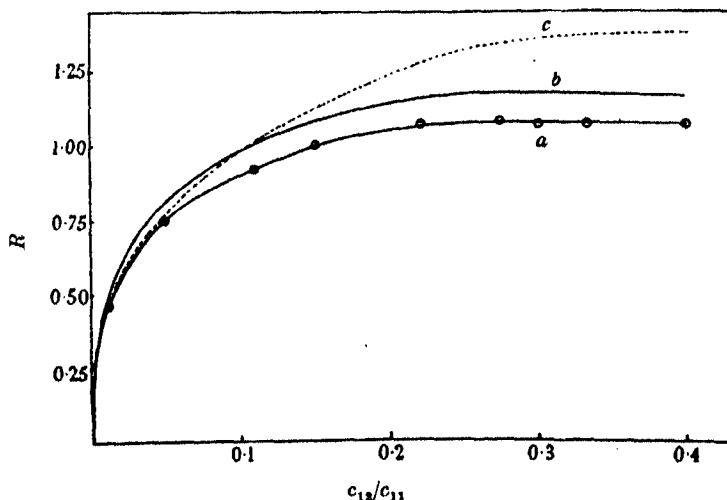
FIGURE 1. The ratio $R = \theta_D(\text{el.})/\theta_D(\text{lat.})$ as a function of c_{12}/c_{11} (or c_{44}/c_{11}): (a) NaCl type and equal masses, (b) CsCl type and equal masses, (c) simple cubic type. The values of R from which the curve is constructed are shown as circles in (a); the same values of c_{12}/c_{11} were used for calculating R in (b) and (c).

of mass difference, as will be clear from (13), is small, e.g. for $m_1/m_2 = 2$ the value of the 'constant' becomes 1·00. It is hence clear that despite the differences between Debye theory and lattice theory, the two theories do give practically the same results as regards this ratio for a number of crystals of the NaCl type. Lattice theory does, however, go further and shows that the ratio tends to zero for very small values of c_{12}/c_{11}; this confirms the result mentioned in the introduction. It should also be

noted that the isotropic case $c_{12}/c_{11} = \frac{1}{3}$ is not exceptional, though the cases where deviation from the Debye theory occur are anisotropic. As regards the continuation of the curve beyond the values calculated, there is evidence that the flat region does not continue. This can be obtained from a set of values of the function $g(c_{12}/c_{11}, c_{44}/c_{11})$ constructed from known data (Blackman 1938). These values indicate that the ratio R will drop to about $0\cdot7$ at $c_{12}/c_{11} = 0\cdot8$; the final value will be zero at $c_{12}/c_{11} = 1$, where the lattice becomes unstable. The portion of the curve for large value of c_{12}/c_{11} is, however, not of great interest, as it is unlikely that the assumptions underlying the calculation of R will be sufficiently accurate.

The rather striking agreement of the lattice theory and Debye theory for the value of R over a limited set of c_{12}/c_{11} values does, however, raise the question, as to whether the agreement, as far as it goes, is a general one, i.e. whether the result is true in the case of other types of forces and crystal structures.

As an example of similar forces to those above combined with a different crystal structure, we can consider the 'body-centred' CsCl type. Using the same assumptions as above in respect of the non-Coulomb forces, one obtains an expression for R which differs from (13) only in the value of the numerical constant, namely,

$$R = \pi \left(\frac{3}{2\pi}\right)^{\frac{1}{3}} \left(\frac{9}{10}\right)^{\frac{1}{3}} \left(\frac{4m_1 m_2}{(m_1 + m_2)^2}\right)^{\frac{1}{3}} \frac{g(c_{12}/c_{11})}{(1 + 2c_{12}/c_{11})^{\frac{1}{3}}}. \tag{14}$$

Curve (b) of figure 1 shows the curve (for equal masses); the constant part of the curve gives a value of $R = 1\cdot16$. In other words, the good agreement of the Debye theory does not survive the change of crystal structure.

The simple cubic lattice with forces of the type introduced by Born & v. Kármán (1912), provides an example of forces different from those of the previous two cases. The details of the calculation have been given above (§ 2 and § 1), and the resulting expression for R is

$$R = \left(\frac{3}{4\pi}\right)^{\frac{1}{3}} \left(\frac{6}{5}\right)^{\frac{1}{3}} \pi g(c_{12}/c_{11}). \tag{15}$$

The resulting curve (curve (c), figure 1) is different in shape from the previous curves and rises to a much larger value (1·40). There is only a small range of values of c_{12}/c_{11}, where R is near the Debye value of unity. The curve will drop for larger values of c_{12}/c_{11}, reaching zero for $c_{12}/c_{11} = 1$.

These examples suffice to show that except in certain special cases there is no exact agreement between θ (el.) and θ_D (sp. ht.); the good agreement obtained in the case of some alkali halides must therefore be considered a fortunate accident.

4. The considerations given above have so far been of a purely theoretical nature. θ_D (el.) and θ_D (sp. ht.) (at high temperatures) have been deduced with the help of the equations for the vibrations of a lattice. From the equations one can, in principle at any rate, obtain the vibrational spectrum; this can also be approximated by a Debye spectrum, and in the resulting theory the ratio R (as calculated in § 3) should be unity. The above considerations are, in this sense, an investigation into the relation of Debye theory and lattice theory.

Before the calculations (for the NaCl type) can be applied to the experimental results quoted in table 1, there are several factors which have to be considered. First, the calculation of the frequencies involved in (2) have been carried out at the absolute zero; and hence it is also assumed implicitly that the elastic and other constants involved in θ_D (el.) are also calculated at the absolute zero; secondly, the calculation assumes $T \gg \theta_D$, whereas the data in table 1 refers to $T \sim \theta_D$. Assuming for the moment that the frequencies were reasonably accurate, this latter point could be looked into by calculating a further term in the expansion of the specific heat as a function of temperature. This calculation has been carried out* and shows that the θ_D value should be only 2 % lower (at $T = \theta_D$) than the asymptotic value. The correction is within the accuracy one would expect for the ionic model and can therefore be omitted. A further point is that the contribution of anharmonic terms to θ_D may be significant even at $T = \theta_D$, whereas the theory assumes harmonic vibrations. In the case of NaCl, the value of C_v at high temperatures has been obtained from the experimental C_p values by Eucken & Dannöhl (1934). By fitting an empirical curve, C_v as a function of T, to these results, one finds that at $T = \theta_D$ the contribution due to these 'extra' terms is certainly smaller than 2 %.

The remaining point to be considered is the most serious one, namely, that the frequencies have been calculated under assumptions strictly true only at the absolute zero. Both θ_D (el.) and θ_D (sp. ht.) are, however, calculated under the same conditions. If it is assumed that the temperature dependence of low-frequency vibrations is the same as of the high-frequency vibrations, then the ratio R will be independent of temperature, i.e. the results of §3 can be applied to the relevant crystals in table 1, as θ (el.) is evaluated there at the temperature $T = \theta_D$. This assumption is a reasonable one, especially as we are dealing with variations which are small; e.g. the variation of θ_D (el.) with temperature in the region considered will be at the most 10 %. In general both θ values tend to decrease with rising temperature, so that it would seem reasonable to estimate the inaccuracy introduced as < 5 %. An error of a few per cent is, however, to be expected in any event, as there is evidence that the ionic model for the NaCl type does not give an exact representation of the forces, e.g. the Cauchy relation $c_{12} = c_{44}$ does not appear to be exactly fulfilled at the absolute zero; Durand (1936) and Rose (1936) give $c_{12}/c_{44} = 0.88$ for NaCl at $T = 0°$ K, and $c_{12}/c_{44} = 0.90$ for KCl at $T = 0°$ K.

Finally, one may compare the value of R obtained from (13) with those quoted in table 1.

TABLE 2

substance	R (exp.)	R (theor.)
NaCl	1·05	1·04
KCl	1·03	1·06
KBr	1·00	1·00

* The details of the calculation will be given elsewhere in a discussion of the bearing of the above results on the specific heat of crystals.

REFERENCES

Blackman, M. 1935*a* *Proc. Roy. Soc.* A, **148**, 384.
Blackman, M. 1935*b* *Proc. Roy. Soc.* A, **149**, 126.
Blackman, M. 1937 *Proc. Roy. Soc.* A, **159**, 416.
Blackman, M. 1938 *Proc. Roy. Soc.* A, **164**, 62.
Born, M. 1920 *Ann. Phys., Lpz.*, **61**, 87.
Born, M. 1923 *Atomtheorie des festen Zustandes.* Leipzig.
Born, M. & v. Kármán, Th. 1912 *Phys. Z.* **13**, 297.
Born, M. & v. Kármán, Th. 1914 *Phys. Z.* **14**, 15.
Debye, P. 1912 *Ann. Phys., Lpz.*, **39**, 789.
Durand, M. 1936 *Phys. Rev.* **50**, 453.
Eucken, A. & Dannöhl, W. 1934 *Z. Elektrochem.* **40**, 789.
Fine, P. C. 1939 *Phys. Rev.* **56**, 355.
Kellermann, E. W. 1940 *Phil. Trans.* A, **238**, 513.
Rose, F. C. 1936 *Phys. Rev.* **49**, 50.
Thirring, H. 1913 *Phys. Z.* **14**, 867.

On the generation of waves on shallow water by wind

By H. BONDI, *Trinity College, Cambridge*

Jeffreys's theory of the formation of water waves is amplified by adding a term for the dissipation in the boundary layer. Considerations of energy lead to an inequality from which values are derived for the minimum wind velocity necessary to excite waves and for the length of the waves thus generated.

The problem of the generation of water waves by wind was first investigated by Lord Kelvin (1871), who examined the stability of the surface of separation of two ideal liquids, both of great vertical extent. Kelvin's theory was restricted to the case of two-dimensional disturbances, but was extended by Jeffreys (1925) to the three-dimensional case. The motion of both liquids is supposed to be irrotational. As is shown in Jeffreys's paper, the predictions of this theory, if applied to the air-water case, entirely disagree with observation.

Jeffreys, who also discusses the equally unsatisfactory hypothesis of skin friction, then evolves the theory of 'sheltering'. The motion of the water is still supposed to be approximately irrotational, but the motion of the air is assumed to be discontinuous. The air current is supposed to separate from the surface at the crest of the wave, striking the windward slope of the next wave some way below its crest. Thus a systematic pressure difference is produced. The 'exposure coefficient' *s* plays a vital part in the theory. It can be determined by observation of the

minimum wind velocity necessary to produce waves, and then the wave-length of the ensuing waves is predicted. Dissipation of energy by viscosity is included in the theory and the agreement with observation is satisfactory.

In a second paper, Jeffreys (1926) extends the theory to the case of water of uniform depth h. The motion is again supposed to be irrotational. Agreement with observation is good for depths of several centimetres but not for smaller depth. For then the wave-length is very much smaller than its value as given by Jeffreys's theory; he has also expressed doubts of the legitimacy of the neglect of some terms in the dissipation. The present paper proposes to discuss the problem mathematically.

As in Jeffreys's papers, first-order effects only are considered, as is indeed logical in the examination of a question of stability. The motion of the water is still supposed to be irrotational near the surface, but the existence of a boundary layer at the bottom is assumed.

In Jeffreys's notation, the velocity potential of the flow outside the boundary layer is

$$\phi = -\frac{\gamma a}{r} \frac{\cosh r(z+h)}{\sinh (rh)} \sin (\gamma t - \kappa x) \cos (\kappa' y), \tag{1}$$

where $r^2 = \kappa^2 + \kappa'^2$. Then the elevation ζ of the surface is given by

$$\zeta = a \cos (\gamma t - \kappa x) \cos (\kappa' y). \tag{2}$$

In the boundary layer $(z+h$ small$)$ I make the usual assumption that w is of the second order of smallness while $\partial/\partial z$ (except for $\partial p/\partial z$) is very large compared with $\partial/\partial x$ and $\partial/\partial y$. Then it is readily seen that

$$\left(\frac{\partial}{\partial t} - \nu \frac{\partial^2}{\partial z^2}\right)(u, v) = -\frac{1}{\rho}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) p, \tag{3}$$

the differential quotients on the right-hand side being taken just outside the layer.

Outside the layer $-\frac{1}{\rho} p = \frac{\partial \phi}{\partial t}$. Also $\left(\frac{\partial \phi}{\partial z}\right)_{z=-h} = 0$.

Hence, neglecting second-order terms,

$$\left(\frac{\partial}{\partial t} - \nu \frac{\partial^2}{\partial z^2}\right)(u, v) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)\left(\frac{\partial \phi}{\partial t}\right)_{z=-h}. \tag{4}$$

Put
$$u = u' + \frac{\partial \phi}{\partial x}, \quad v = v' + \frac{\partial \phi}{\partial y}. \tag{5}$$

Then, to the same degree of approximation,

$$\left(\frac{\partial}{\partial t} - \nu \frac{\partial^2}{\partial z^2}\right)(u', v') = 0. \tag{6}$$

As $z \to \infty$, $(u', v') \to (0, 0)$, and, at $z = -h$, $(u', v') = -\left(\dfrac{\partial}{\partial x}, \dfrac{\partial}{\partial y}\right)\phi$. The integrals are easily found, viz.

$$u' = -\frac{\gamma a \kappa}{r \sinh(rh)} \exp\left[-\sqrt{\left(\frac{\gamma}{2\nu}\right)}(z+h)\right]\cos\left[\gamma t - \kappa x - \sqrt{\left(\frac{\gamma}{2\nu}\right)}(z+h)\right]\cos(\kappa' y), \quad (7)$$

$$v' = -\frac{\gamma a \kappa'}{r \sinh(rh)} \exp\left[-\sqrt{\left(\frac{\gamma}{2\nu}\right)}(z+h)\right]\sin\left[\gamma t - \kappa x - \sqrt{\left(\frac{\gamma}{2\nu}\right)}(z+h)\right]\sin(\kappa' y). \quad (8)$$

According to Lamb (1906, p. 541) the rate of dissipation of energy is given by

$$4\mu \iiint (\xi^2 + \eta^2 + \zeta^2)\,d\tau + \mu \iint \frac{\partial(u^2 + v^2 + w^2)}{\partial n}\,dS - 4\mu \iint \begin{vmatrix} l & m & n \\ u & v & w \\ \xi & \eta & \zeta \end{vmatrix} dS, \quad (9)$$

where $\xi = \dfrac{1}{2}\left(\dfrac{\partial w}{\partial y} - \dfrac{\partial v}{\partial z}\right)$, etc., dn denotes an element of the outward normal, and l, m, n are its direction cosines. The second integral has been considered by Jeffreys (1926, p. 243), who obtains the expression

$$\mu r \gamma^2 a^2 \coth(rh) \quad (10)$$

for the average rate of dissipation of energy (due to this term) in a column of unit cross-section. The contribution of the third term is negligible, since at the bottom $u = v = w = 0$, while on the surface the vorticity is very small.

Consider now the volume integral. All the differential quotients occurring are small except for $\partial u/\partial z$ and $\partial v/\partial z$. Hence, to the usual degree of approximation,

$$4(\xi^2 + \eta^2 + \zeta^2)_{\text{mean}} = \left[\left(\frac{\partial u}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2\right]_{\text{mean}} = \frac{\gamma^3 a^2}{4\nu \sinh^2(rh)} \exp\left[-\sqrt{\left(\frac{2\gamma}{\nu}\right)}(z+h)\right], \quad (11)$$

where mean values have been taken over x, y, and t.

Hence the contribution of this integral is

$$\frac{\rho \gamma^2 a^2}{4 \sinh^2(rh)} \sqrt{\left(\frac{\gamma\nu}{2}\right)} \quad (12)$$

if $\exp[-\sqrt{(2\gamma/\nu)}\,h]$ is neglected.

Now according to Jeffreys (1926, p. 244) the mean value of the rate at which the wind pressure is doing work is

$$\tfrac{1}{2}s\rho'(V - c)^2 \gamma \kappa a^2, \quad (13)$$

where $c = \gamma/\kappa$ is the wave velocity and V is the wind velocity. Hence the wave will grow if

$$\tfrac{1}{2}s\rho'(V - c)^2 \gamma \kappa a^2 > \nu \rho r \gamma^2 a^2 \coth(rh) + \frac{\rho \gamma^2 a^2}{4 \sinh^2(rh)} \sqrt{\left(\frac{\gamma\nu}{2}\right)}. \quad (14)$$

This can be written

$$\frac{(V-c)^2}{c} > \frac{4\rho}{s\rho'} \left\{ \nu r \coth(rh) + \sqrt{\left(\frac{\nu\gamma}{2}\right) \frac{1}{4\sinh^2(rh)}} \right\}. \tag{15}$$

If $\kappa' = 0$, (10), (11), (12) and (13) will all be doubled, but this will clearly leave (14) and (15) unchanged.

Again, according to Jeffreys (1926, p. 242)

$$\kappa^2 c^2 = \gamma^2 = \left[g\frac{\rho-\rho'}{\rho} + T'r^2 \right] r \tanh(rh). \tag{16}$$

With the exception of the second term on the right-hand side (14) and (15) are identical with Jeffreys's condition (1926, p. 244). The new term, which decreases rapidly as rh increases, is large for small rh.

If now r is kept constant but κ is varied, the right-hand side of (15) remains unchanged. The left-hand side is, for $V > c$, a decreasing function of c and hence an increasing function of κ. Thus the inequality is most likely to be satisfied if κ attains its maximum value, i.e. $\kappa = r$, $\kappa' = 0$. Hence, just as in Jeffreys's case, the easiest waves to produce are two-dimensional.

In this case (15) and (16) become

$$\begin{aligned} \frac{(V-c)^2}{c} &> \frac{4\rho}{s\rho'} \left\{ \nu\kappa \coth(\kappa h) + \sqrt{\left(\frac{\nu\kappa c}{2}\right) \frac{1}{4\sinh^2(\kappa h)}} \right\}, \\ c^2 &= \left(\frac{g}{\kappa}\frac{\rho-\rho'}{\rho} + T'\kappa \right) \tanh(\kappa h). \end{aligned} \tag{17}$$

The validity of this inequality depends on the correctness of our assumptions.
The application of the boundary layer theory is justified if

$$\frac{\partial}{\partial z} \gg \frac{\partial}{\partial x} \quad \text{at} \quad z = -h, \quad \text{i.e.} \quad c \gg \nu\kappa. \tag{18}$$

Also $\exp[-\sqrt{(2\gamma/\nu)}\,h]$ may be neglected, if

$$1 \ll \sqrt{\left(\frac{2\gamma}{\nu}\right)}h, \quad \text{i.e.} \quad h^2\kappa c \gg \nu. \tag{19}$$

By considering first-order terms only, a necessary but possibly not a sufficient condition for the formation of noticeable waves is found in (17). For short waves in particular V may have to be somewhat greater than is suggested by (17). For then $u(\partial u/\partial x)$ might not be small compared with $\partial u/\partial t$ for noticeable values of a.

The minimum values of $V(V_m)$ satisfying (17) were calculated together with the corresponding values of $\kappa(\kappa_m)$ for various depths h (see below).

The values of the constants were taken to be

$$g\frac{\rho-\rho'}{\rho} = 980\,\text{cm./sec.}^2, \quad T' = 73\,\text{cm.}^3/\text{sec.}^2, \quad \nu = 0.018\,\text{cm.}^2/\text{sec.}, \quad \frac{4\rho}{s\rho'} = 12321.$$

This makes s approximately 0.25, a value suggested by Jeffreys's experiment (1925, p. 199).

For smaller values of ν and $4\rho/s\rho'$, V_m will be smaller, while κ_m will be closer to the value of κ corresponding to minimum wave velocity.

As $h \to 0$, $\kappa_m \to \infty$, $\kappa_m h \to 0$, and $c_m \to 0$. Hence, for very small h, (18) and (19) are no longer satisfied. Indeed for $h < 0.5$ cm. the (κ_m, h) curve has a kink, while it is very regular elsewhere. In fact, it is not suggested that (17) has much physical significance for $h < 0.5$ cm. For reasons suggested above V may have to be somewhat larger than V_m in order to produce noticeable waves.

Otherwise the theory seems to agree well with observational evidence (as far as this goes) in that λ_m decreases strongly as h decreases, while V_m increases at the same time.

In conclusion, I wish to express my gratitude to Dr Jeffreys for suggesting the subject to me and giving me his views on it.

<div align="center">TABLE</div>

$$V = c + \sqrt{\left[\frac{4\rho}{s\rho'} c \left\{\frac{\nu\kappa}{\tanh(\kappa h)} + \sqrt{\left(\frac{\nu\kappa c}{2}\right)} \frac{1}{4\sinh^2(\kappa h)}\right\}\right]} = f(\kappa),$$

$$c^2 = \left(\frac{1}{\kappa} g \frac{\rho - \rho'}{\rho} + T'\kappa\right) \tanh(\kappa h),$$

$$f'(\kappa_m) = 0, \quad V_m = f(\kappa_m), \quad \kappa_m \lambda_m = 2\pi.$$

h (cm.)	κ_m (cm.$^{-1}$)	λ_m (cm.)	V_m (cm./sec.)	h (cm.)	κ_m (cm.$^{-1}$)	λ_m (cm.)	V_m (cm./sec.)
0.25	5.8	1.08	251.5	3	1.11	5.65	120.0
0.5	3.85	1.63	182.9	4	0.90	7.00	117.2
1	2.48	2.56	145.5	5	0.783	8.02	115.9
1.25	2.07	3.03	137.6	8	0.61	10.3	114.9
2	1.43	4.30	125.3	∞	0.598	10.48	114.8

$$\sqrt{\left(\frac{4\rho}{s\rho'}\right)} = 111, \quad g\frac{\rho - \rho'}{\rho} = 980 \text{ cm./sec.}^2, \quad T' = 73 \text{ cm.}^3/\text{sec.}^2, \quad \nu = 0.018 \text{ cm.}^2/\text{sec.}$$

REFERENCES

Jeffreys, H. 1925 *Proc. Roy. Soc.* A, **107**, 189–206.
Jeffreys, H. 1926 *Proc. Roy. Soc.* A, **110**, 241–247.
Kelvin, Lord 1871 *Phil. Mag.* (4), **42**, 368–370. See also Lamb, H. 1906 *Hydrodynamics*, p. 439.
Lamb, H. 1906 *Hydrodynamics*.

A stress-strain curve for the atomic lattice of mild steel in compression

BY S. L. SMITH, D.Sc., F.C.G.I., *Engineering Department*
AND W. A. WOOD, D.Sc., *Physics Department*

The National Physical Laboratory, Teddington, Middlesex

A stress-strain curve has been obtained for the atomic lattice of mild steel subjected to compression. A set of atomic planes is selected of which the spacing is practically perpendicular to the direction of the stress, and the change in spacing is measured as the magnitude of the applied stress is systematically varied. The behaviour of the lattice is compared with the corresponding stress-strain relation for the external dimensions in the compression test, and also with the lattice stress-strain curve previously obtained for the same material when subjected to tensile stress. Other experiments are described on the behaviour of the lattice of pure iron in compression. It had been previously shown that at the external yield in tension, the atomic spacing exhibited an abrupt change which remained indefinitely on removal of the stress; the effect was interpreted as a lattice yield point. The present work establishes that the lattice possesses a yield point also in compression, again marking the onset of a permanent lattice strain. The direction of this strain, however, is opposite to that found in tension, and the magnitude increases systematically with the applied stress. The experiments on the pure iron show that under extreme deformation the permanent lattice strain tends to a limit and that with continued deformation partial recovery from the strain may occur. The results suggest that the mechanics of the metallic lattice involve the principle that, after the lattice yield point, in a given direction the lattice systematically assumes a permanent strain in such a sense as to oppose the elastic strain induced by the applied stress.

INTRODUCTION

Investigations have been carried out recently on the deformation of the atomic lattice of metals under systematically applied stresses (Smith & Wood 1940, 1941 a, 1941 b). The object has been the elucidation of the laws relating the displacement of the atoms to the external stress conditions in the metal.

So far, the experiments have been concerned with *tensile* stress, and the displacements measured have been displacements in the direction perpendicular to the tensile stress. The main experiments described in the present paper represent a further stage by indicating the behaviour of the lattice under *compressive* stress. The atomic displacements have again been measured in the direction perpendicular to the stress, partly for convenience of technique but also to provide results which could be contrasted with those obtained in tension. These first experiments with compression have been made chiefly on mild steel, since this material is amongst those of which the behaviour under tensile stress has already been determined, but some experiments on pure iron are also described. Besides providing confirmation of the results on the mild steel, and showing the influence of the carbon in steel, these latter experiments were used to bring out special aspects in the deformation of the metallic lattice.

[72]

The displacements perpendicular to the applied stress have been found to be of such a nature that the displacement actually along the line of the stress presents a particularly intriguing problem. Measurements in that direction, however, involve modifications of apparatus, which are in hand. But in the meantime some indications have been obtained of the way changes in the line of the stress are likely to go, and these are referred to in this paper since they influence the interpretation of the main experiments on compression.

<center>EXPERIMENTAL</center>

The experimental arrangements were essentially the same as those used in the previous work on tension, the external strain exhibited by the test piece being measured by extensometer as known stresses were applied, and the lattice strain by changes in diameter of a sensitive X-ray diffraction ring formed by back-reflexion of an incident beam which was perpendicular to the length of the specimen, and therefore to the stress direction. Instead of a small testing machine, however, a standard 10-ton testing machine was utilized, and the X-ray tube built up alongside in a position suitable for taking the back-reflexion photographs. The large-scale machine was necessary for the compression tests because of the greater cross-section of specimens for such tests, and the greater loads thereby involved. The specimens were solid cylinders 1·5 in. long and 0·4 in. diameter, or 0·75 in. long and 0·358 in. diameter, the shorter specimens being for use with particularly heavy loads to obviate buckling. The steel was received in the normalized condition, and was further heated for approximately ½ hr. at 600° C in vacuo to remove machining effects. Specimens were used only after preliminary X-ray examination had shown reflexion spots which were sharp enough to indicate freedom from residual strains in the individual grains.

The use of the larger machine did not allow of oscillation of the specimen during the X-ray exposure. It is well known that at the large reflexion angles used in the back-reflexion technique the grains reflect the incident beam for a few seconds of arc about the peak reflexion angle. It is desirable to oscillate the specimen at least through that range in order to obtain the peak, which then becomes independent of the setting of the grain. Also, an appropriate oscillation brings more grains into a reflecting position; the reflexion spots then tend to coalesce into a continuous ring, which facilitates measurements of diameter. These difficulties were largely overcome by using as broad an incident beam as was practicable. A point on the specimen then received rays from a small convergent cone, which was equivalent to oscillating the specimen in a parallel beam. Also the increase in area under the X-ray beam resulted in more reflecting grains. Conditions were finally obtained which gave a diffraction ring sufficiently continuous for the required accuracy of measurement. In order, however, to make sure of the changes in diameter remaining after a particular stress had been removed, a point which has assumed special importance in this work, confirmatory experiments were made. For this purpose, some

specimens were loaded in the machine to selected stresses, and, after unloading, were immediately removed to an X-ray spectrometer where they could be oscillated and examined under precise and well-tried conditions.

The use of an incident beam directed at right angles to the length of the test specimen was dictated by the necessity of avoiding the plattens fixing the ends of the specimen in the machine. The atomic planes from which back-reflexions were recorded were therefore planes lying approximately parallel to the length of the specimen, and with an interplanar spacing approximately perpendicular to that direction, and therefore to the applied stress. With cobalt $K\alpha$ wave-length as incident radiation, a reflexion is given by the (310) planes at $\theta \sim 81°$, where the sensitivity to changes of spacing is high, since the Bragg relation $2d \sin \theta = \lambda$ gives $\delta\theta = -\tan \theta(\delta d/d)$. The measurements therefore refer to this spacing, and to those grains in which the (310) spacing makes the angle of $81°$ with the line of the stress, since such grains only are in a position which satisfies the Bragg condition for reflexion. The behaviour of this spacing, and of the group of grains involved, will be representative of the behaviour of other spacings which lie practically perpendicular to the stress direction.

External stress-strain curve

Although the main interest of the investigation lay in the properties of the lattice, it is convenient to record first the stress-strain curve for the external dimensions. This is shown by figure 1. The strain there plotted is the lateral expansion in diameter divided by the original diameter; the stress is the actual compressive stress applied along the length of the test piece, given by dividing the total load by the measured area of cross-section reached at that load. The curve shows the sharp onset of plastic strain at the yield point which is a particular characteristic of mild steel, and which forms a point of special interest when contrasted with the behaviour of the lattice at the same stress.

Previous experience had shown that the properties of the lattice were best brought out by unloading the test piece after each increase in applied stress, and obtaining an X-ray photograph at each stress and also after each stress had been removed. Figure 1 refers to such a procedure. Thus the specimen was X-rayed at a point such as P_4, and then when unloaded at the point M_4; the stress was increased to the next point P_5
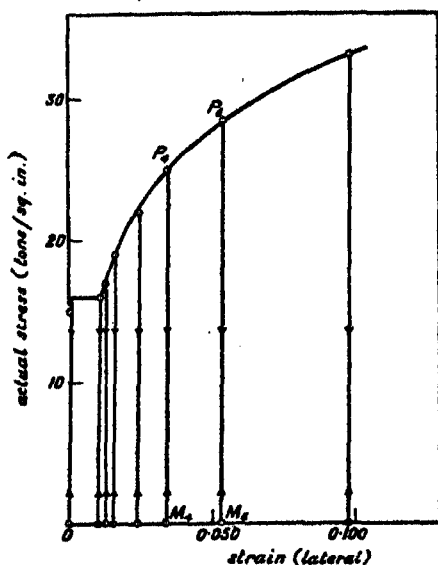

FIGURE 1. External stress-strain curve for mild steel (compression). The strain refers to expansion in diameter divided by original diameter.

and photographs again obtained under stress and after unloading, in this case to M_5. The stress-strain curve thus followed the path, $\ldots P_4 M_4 P_4 P_5 M_5 P_5, \ldots$. This procedure gave in one test, for all practical purposes, the variation in atomic spacing with continued progressive loading, and also the extent of any permanent change in spacing produced at each stress and remaining after that stress was removed.

The mild steel (0·1 % C) was from the same batch as that used in the previous work on tensile stress (Smith & Wood 1941 b). The principal mechanical properties were as follows:

	tension tons/sq. in.	compression tons/sq. in.
Young's modulus	$1·3 \times 10^4$	$1·3 \times 10^4$
yield stress	16·8	16·0
ultimate stress	27·7	——
Poisson's ratio		0·28

Lattice stress-strain curve

The variations in (310) atomic spacing, as measured in the direction approximately perpendicular to the stress direction, are best set out by the stress-strain curves shown in figure 2. The strain there plotted is the percentage change from the initial value of the (310) spacing; expansion is plotted to the left of the origin and contraction to the right, as in the previous work on tension. In that work, the stress was plotted as the positive ordinate: in the present tests the compression stress has been plotted below the origin as negative ordinate. If figure 2 is then compared with the corresponding curve for tension (reproduced here in figure 3), the effect on the (310) spacing of substituting tension for compression becomes at once evident.

The compression curve, like the tensile curve, indicates three phases in the behaviour of the lattice.

The *first phase* is the behaviour before external yield takes place. There the (310) spacing expands perpendicularly to the compressive stress, and returns to its original value when the stress is removed. The spacing obeys Hooke's law and follows the related expansion and recovery of the external cross-section of the specimen. The point of this phase is that, within limits of measurement, no residual change remains in the lattice after removal of the applied stress.

The *second phase* concerns the behaviour of the lattice when, externally, the specimen is brought to the yield point. Up to the yield stress, as indicated in the first phase, the atomic spacing under consideration expands in proportion to the applied stress. When the yield stress is reached, this expansion ceases. The spacing exhibits an abrupt *contraction*. This contraction is the more remarkable because the corresponding external dimension (the diameter of the specimen perpendicular to the stress) exhibits an equally abrupt *expansion*, as the onset of plastic deformation occurs. The feature of this contraction in spacing is that it is a permanent strain, since, as indicated by figure 2, it remains when the stress is removed. The extent of the change is shown in that figure by OM_1. The characteristic of this second phase

therefore is the appearance in the lattice of a permanent and new type of strain associated with a mechanism of deformation which is fundamentally different from the simple elastic strain of the preceding phase.

The *third phase* occurs as the applied stress is continued beyond the yield. Here there are two features. The first relates to the magnitude of the atomic spacing whilst under stress. It is seen from figure 2 that, after a little irregularity in the immediate neighbourhood of the yield itself, the spacing settles down to an
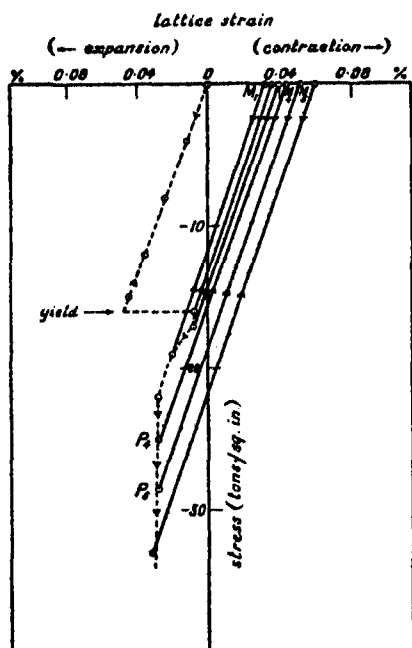
FIGURE 2. Lattice stress-strain curve for mild steel in compression. The strain refers to the percentage change in (310) spacing in the direction perpendicular to the stress.
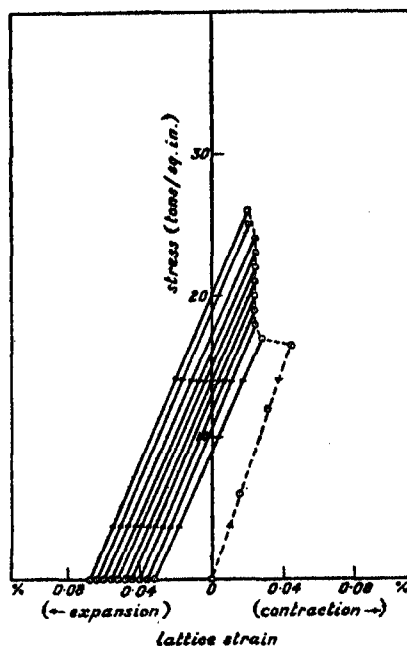
FIGURE 3. Lattice stress-strain curve for mild steel in tension. The strain refers to the percentage change in (310) spacing in the direction perpendicular to the stress.

approximately constant value, despite the progressive increase in stress and the continued expansion in the related external diameter of the specimen. The constant range is shown by $P_4 P_5 \ldots$. The second feature relates to the behaviour of the spacing when the stress is removed. Then, the spacing is left with a permanent contraction which increases regularly in magnitude with increases in the stress from which unloading takes place. Thus, after unloading from P_4, the spacing exhibits the contraction corresponding to OM_4; after unloading from the higher stress P_5, the contraction increases to OM_5, and so on. The characteristic of this third phase therefore is the regular increase with stress of the permanent residual lattice strain and the superposition of this residual strain on the reversible elastic

expansion induced by the applied stress. The two types of strain are so related that their algebraic sum results in an approximately constant lattice spacing at stresses beyond the yield.

These changes may now be compared with the behaviour of the same lattice spacing measured in the same direction, perpendicular to the stress, but under tensile stress. The corresponding tensile lattice stress-strain curve is shown in figure 3. When this curve is contrasted with figure 2, it will be seen that the substitution of compression for tension reverses the sign of the change in spacing in each of the three phases described above. In the elastic range, referred to as the first phase, such a reversal was to be expected. But there was nothing to indicate that the permanent set of the lattice would also change from the irreversible expansion in tension to the irreversible contraction observed in compression, especially since in both cases the direction of the effect is unexpectedly in opposition to the corresponding changes in external dimensions.

It will also be seen from figures 2 and 3 that the magnitude of the new effect is not a second-order effect but is comparable with the changes in dimensions associated with the elastic deformation of metals. The most interesting quantitative data are summarized below. The slight differences shown in this summary between the measurements in the compression experiments and those in the tensile tests probably arise because, although the same material was used, the specimens for the two sets of experiments were of different shape, and because the processes of deformation in the standard compression and tensile tests are not strictly complementary.

Quantitative data

(1) Over the range of variation observed, the changes in diameter of the diffraction ring were proportional to the changes in atomic spacing. A difference in diameter of 1 mm. under the experimental conditions employed corresponded to a difference in atomic spacing of 0·04 %. The diameters of the rings could be compared accurately to the nearest 0·1 mm., which was sufficient for the purpose of the investigation.

(2) Writing + for lateral expansion, the elastic change in (310) spacing per unit increase in stress (ton/sq. in.), at 81° to the stress direction, was: compression, $+0\cdot0029\ \%$; tension, $-0\cdot0027\ \%$. If these values are divided by cos 9° to obtain an indication of the change actually at right angles to the stress direction, they become $+0\cdot0029_4$ and $-0\cdot0027_4$ %. The values at 81° as measured may therefore be taken to be the same, for all practical purposes, as those at right angles to the stress.

(3) It is of interest to compare the foregoing changes with the corresponding changes in external diameter of the test specimens. The elastic change in diameter perpendicular to the stress per unit increase in stress was: compression, $+0\cdot0023\ \%$; tension, $-0\cdot0022\ \%$. These values have been obtained from measurements of the longitudinal strain and measurements of Poisson's ratio. It will be noted that,

whilst of the same order of magnitude, they are not identical with the related lattice changes. Some difference is, however, to be expected, since the external measurements average out the changes in a number of (h, k, l) lattice spacings of which the (310), though representative, is only one. The departure of the (310) values from the average indicates the anisotropy of the lattice.

(4) The extent of the permanent change in the (310) spacing which took place at the yield point, OM_1 in figure 2, is also of interest in showing that the effect there observed is quite large compared with the more normal elastic lattice strains, and, moveover, that it is of the same order in tension as compression. The change was: compression, -0.03%; tension, $+0.03\%$.

(5) Finally, it is interesting to note the marked extent of the new permanent lattice-strain effect which can be reached at the higher ranges of the stresses employed in these tests. Thus at 33 tons/sq. in. in compression and at the ultimate stress in tension, 28 tons/sq. in., the values reached were as much as: compression, -0.06%; tension, $+0.07\%$.

Confirmatory experiments on iron (purity 99.95 %)

It had previously been shown that the iron lattice in tension behaved similarly to mild steel in tension (Smith & Wood 1941 b). The only difference was one of degree: the permanent lattice-strain effect in iron, though well marked, does not reach the magnitude observed in mild steel; also the strain does not set in as abruptly at the yield point, an observation corresponding to the less abrupt external yield exhibited by the pure metal.

The main object of the present confirmatory experiments was to test whether the new effect now found in mild steel under compression existed also in the pure iron under compression, and in the same direction. A positive result would also confirm that the effects were not associated with the carbide formations inherent in a steel but were fundamental properties of the metallic lattice itself.

A further aim of the experiments on pure iron was to take advantage of the suitability of this material for measurements on the permanent lattice strain after particularly heavy deformation of the test specimens. This suitability arises because the lower limiting crystallite size for pure iron is larger than for steel; as a result, the diffraction rings remain sharper under comparable deformations of the specimens, owing to the lesser degree of diffusion introduced by the Scherrer effect (Wood 1939, 1940). The tests with heavy deformation are, moreover, more easily carried out in compression than tension, since, with the latter, the specimens neck and quickly fracture after the ultimate stress is reached, whilst, with compression, the limit to precise measurements is set in practice only by irregular deformation and by the specimen becoming barrel-shaped.

Both objects could be served by confining the measurements to the magnitude of the permanent lattice strain remaining in the specimens after removal of selected values of the applied stress. Therefore after a particular stress had been removed, the specimen under test was taken from the machine and examined in a back-

reflexion spectrometer under standard conditions. It was then replaced in the machine for the next higher stress and the process repeated. The time for which the specimen was subjected to each stress was about 5 min.; the interval between successive stresses whilst the specimen was being removed and X-rayed was approximately $\frac{1}{2}$ hr. It was known from other experiments that the permanent lattice-strain effect remained normally unchanged at room temperature for an indefinite period; the intervals of resting, therefore, were not likely to affect the main course of the results. Moreover, arrangements were made to ensure that the X-ray examination related throughout to the same points on the specimen. Other conditions were similar to those for the tests on the mild steel.

The results are depicted in figure 4, where the lattice strain remaining after each stress is plotted against that stress. As before, the strain refers to the percentage change from the original value of the (310) spacing, measured in the direction at right angles to the stress. The stress is the actual stress, given by the load divided by the area of cross-section of the specimen at each stress. The observed external plastic deformation, as given by the change in diameter divided by the original diameter of the test piece, is also given for reference on the same diagram.

The measurements shown in figure 4 amply confirm the first object of the tests by showing that, in the direction normal to the stress, the (310) spacing *contracts* permanently as soon as a stress beyond the yield point is reached, just as with the mild steel, and that in the early stages the magnitude of the effect increases proportionately with the applied stress. The new effect is therefore a mechanical property of the iron lattice and has nothing to do, primarily, with the carbide formation of a steel. At the same time, the influence of the carbon

FIGURE 4. Relation between permanent lattice strain and applied stress for pure iron in compression. The strain refers to the percentage change in (310) spacing in the direction perpendicular to the stress. The permanent strain at each stress is revealed when that stress is removed.

is of interest: it apparently confers upon the iron lattice a capacity for exhibiting a greater permanent lattice strain consistent with stability.

The shape of the curve in figure 4 for the higher stresses also gives information on the second object of the tests, the behaviour of the lattice after heavy deformation. It will be seen that the rate of increase of the permanent strain effect tends to fall off and reach a limiting value. It is reasonable that some limit should exist, otherwise it would be difficult so see how the lattice could continually preserve its stability in view of the magnitude of the effect observed. The measurements
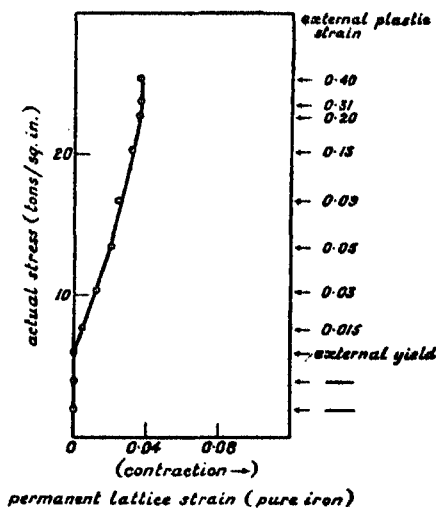
recorded in figure 4 went as far as they could be extended before the accuracy was affected by the development of irregularities in the shape of the specimen under test. Further measurements were, however, made and, though less accurate quantitatively, they were of interest in that they indicated that the effect of still further continued deformation caused uneven variations in the magnitude of the lattice strain. Thus, the strain at different points round the circumference of the specimen, whilst never exceeding the limit indicated by figure 4, ceased to be always the same, and the strain at a particular point appeared occasionally to undergo partial recovery to a less value. The conclusion was made that if the deformation was pushed beyond the magnitude corresponding to the limiting value shown by figure 4, the lattice definitely became unstable, and some form of structural change occurred by which part of the lattice strain was released; it might be suggested that the limiting value represented an equilibrium between this tendency of the lattice to exhibit partial recovery and the capacity for taking up further strain. This conclusion would be in keeping with other work (Wood 1939) on the lattice strain in metals heavily deformed by cold rolling, where the existence of the process of partial recovery was definitely established, and would form a link between the effects of heavy deformation and the more systematic deformation occurring in the ranges of stress normally used in the standard tensile and compression tests on metals.

INTERPRETATION OF RESULTS

The chief contribution to the problem of the deformation of metals of this and the previous related investigations lies in establishing quantitative relations between applied stress and the displacement of the atoms in the direction perpendicular to that stress. From these relations, certain conclusions can be drawn regarding the mechanical properties of the metallic lattice. Of these perhaps the most important, from both technical and theoretical viewpoints, is the observation that the behaviour of the atoms does not follow solely a simple elastic displacement from the unstressed condition to fracture which previously, in the absence of experimental data, has been tacitly assumed. In fact, it has been shown that equally important is a second process, which differs fundamentally in being irreversible, and which corresponds to a marked permanent absorption of energy by the lattice. This irreversible strain is superposed on the elastic strain with the result that the effective lattice stress-strain curve has little obvious relation either to Hooke's law or to the external deformation exhibited by the metal. Moreover, in the range covered by normal tensile and compression tests, the irreversible strain behaves in such a manner that it may be said to constitute a systematic property of the metallic lattice; also, the strain appears at a particular stress in such a manner, especially in mild steel, as to indicate that the lattice itself possesses the property of a 'yield point'.

The striking and unexpected characteristic of the irreversible strain effect, as observed in the direction perpendicular to the applied stress, is that it appears to

take the opposite direction to the elastic strain. Thus, in the compression tests, the atomic spacing under examination expanded in the elastic range as stress was applied, following the external expansion in diameter of the specimen; beyond the elastic range the spacing still expanded to some extent under stress, but when the stress was removed the spacing then exhibited a permanent *contraction* on the original unstressed value. This contraction, representing the irreversible fraction of the total change in spacing, was thus opposite in direction to the elastic part of the strain caused by the application of the stress: the actual displacement of the atoms under stress was therefore the algebraic sum of two oppositely directed processes.

A consideration of this characteristic in the tensile and compression measurements perpendicular to the applied stress then appears to justify the statement of the following principle: *After the lattice yield point, the metallic lattice systematically assumes a permanent strain in such a sense as to oppose the elastic strain induced by the applied stress, and of a magnitude which increases with the magnitude of the stress.*

The foregoing principle should play a fundamental part in the explanation of one of the chief properties of a metal, that of strain-hardening. This property is bound up with the shape of the external stress-strain relationship of a metal. Thus, if reference is made to figure 1, the compression stress-strain curve for mild steel, it is seen that after the plastic strain OM_4, the material exhibits the apparent elastic range corresponding to M_4P_4, which is greater than the primitive elastic range preceding the yield point. If, by increasing the stress, the plastic strain is extended from OM_4 to OM_5, then the apparent elastic range is increased from M_4P_4 to M_5P_5; the effect of increased stressing and increased plastic deformation is thus to enhance the apparent elastic range exhibited by the material, which is then said to be strain-hardened. This property may now be examined in the light of the stress-strain curve for the atomic lattice. Before curves of this type were available, it might have been thought that, under applied stress, the atoms were pulled from their normal positions of equilibrium in the metallic lattice, returning to those positions on release of the stress, and that the apparent increase in external elastic range during the strain-hardening referred to above was due merely to an increase in the extent of the displacement of the atoms with application of greater and greater stress, the atoms again returning to their original positions on removal of the stress. The atomic stress-strain curve of figure 2, which corresponds to the external curve of figure 1, shows that this commonly assumed picture is quite wrong; the displacement of the atoms from their original positions does not proceed continuously with increasing stress, but, in fact, quickly reached a limit; also, after the yield, when strain-hardening occurs, the atoms do not return to their initial positions when the applied stress is removed. It becomes necessary therefore to account for the enhanced external elastic range on quite other grounds. The true position is at once evident on reference to figure 2. There the ranges of atomic displacements M_4P_4 and M_5P_5 correspond to the external elastic ranges M_4P_4 and M_5P_5 in figure 1. It will be seen that the increase in the atomic displacement range

of M_4P_4 to M_5P_5 is rendered possible only because of the increase in the permanent lattice strain from OM_4 to OM_5. The conclusion appears to be justified therefore that the enhanced elastic range does not arise because of the simple elastic properties of the metallic lattice, but because of the capacity of the lattice for assuming a permanent strain in the opposite direction to the elastic displacement induced by the applied stress.

Considerations such as the foregoing, based on experimental measurements of atomic displacements perpendicular to the applied stress, emphasize the importance of knowing the displacements actually in the direction of the stress. Some preliminary experiments on this problem have been undertaken. At first it was thought that information regarding the new permanent lattice-strain effect parallel to the stress direction might be obtained by cutting a tensile or compression test piece into two and making X-ray measurements perpendicularly to the exposed cross-sections. Experiments on those lines were carried out, but after some time it was found that the results were not consistent. The variations were traced to the effect of cutting the specimens, which, amongst other things, caused a release and redistribution of the internal lattice strain in the neighbourhood of the cut or fracture. It is therefore essential for the solution of this problem to perform the necessary experiments on tensile or compression specimens preferably whilst they are in the testing machine or, at any rate, without interfering with the structure. Experimental arrangements for this purpose are in hand. In the meantime, measurements have been made on tensile specimens stretched in the machine and then removed to an X-ray spectrometer where they have been photographed stationary, with the incident X-ray beam at various angles to the length of the specimen along which the stress had been applied. In this way a direct attempt was made to measure the extent of the permanent lattice-strain effect in various directions. Perpendicular to the stress direction, this strain in the tensile specimens is an *expansion*, as shown for the (310) spacing in figure 3, since, in accordance with the above-mentioned principle, it takes the direction opposite to the elastic *contraction* of the cross-section as the specimen is stretched. It was, however, found that at 45° to the stress direction, the expansion had diminished in extent and the spacing was practically equal to the normal for the (310) planes. At 30° and 15° to the stress direction, the (310) spacing was below normal, indicating that the permanent lattice strain was becoming a contraction. It was not practicable, for reasons of technique, usefully to approach the stress direction more closely, but the measurements already obtained sufficed to indicate that, in a tensile specimen *which was not subjected to any cutting or sectioning*, the permanent expansion perpendicular to the stress became a permanent contraction actually along the stress direction. It appears likely that the principle stated above for changes at right angles to the stress will hold for the changes in the lattice parallel to the stress, although the point will only definitely be proved when the foregoing exploratory tests are completed by tests on a specimen actually under applied stress.

The work described above has been carried out as part of the research programme of The National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES

Smith, S. L. & Wood, W. A. 1940 *Nature, Lond.*, **146**, 400.
Smith, S. L. & Wood, W. A. 1941*a* *Proc. Roy. Soc.* A, **178**, 93.
Smith, S. L. & Wood, W. A. 1941*b* *Proc. Roy. Soc.* A, (in the Press).
Wood, W. A. 1939 *Proc. Roy. Soc.* A, **172**, 231.
Wood, W. A. 1940 *Proc. Phys. Soc.* **52**, 110.

Thermal insulation at very low temperatures

By A. H. COOKE AND R. A. HULL

Clarendon Laboratory, Oxford

Various methods of insulation have been investigated at temperatures between 0·2 and 1° K. A simple suspension of artificial silk fibre reduces the heat flow to 5 ergs/min.; 'guard rings' at the same low temperature reduce the heat flow to 1 erg/min. The origin of the heat flow in various experimental arrangements is considered in its bearing on the technique of the magnetic method of cooling.

1. INTRODUCTION

In the region of temperature attainable by the magnetic method, the problem of thermal insulation is somewhat different from that at higher temperatures. The vapour pressure of all gases is extremely low, and the effect of radiation is negligible; the thermal conductivity of amorphous and polycrystalline solids, including metals, is very low (though there is theoretical and experimental evidence that for very pure crystals free from strain the conductivity would increase at very low temperatures). Generally speaking therefore thermal insulation is good in this region. The investigation of the subject is, however, important, not only because of the intrinsic interest of the mechanism of heat transfer in this new range of temperature, but also from the practical point of view. For instance, the insulation is a determining factor for the limits of low temperature attainable (see Simon 1939, where it is pointed out that the good insulation mentioned later in this paper would permit the attainment of the still lower temperatures associated with the nuclear magnetic moments). It also determines the lowest amount of energy that can be measured calorimetrically; with the present insulation it would already be possible to measure directly comparatively weak radioactive changes, e.g. the ordinary uranium disintegrations or the disintegrations of artificial radioactive substances.

Furthermore, the heat flow due to lack of insulation is made more deleterious at these very low temperatures because of the low thermal conductivities which obtain there. At any temperature this heat flow takes place primarily to somewhere on the outer surface of the sample; but whereas at higher temperatures the sample will remain almost uniform in temperature, with low thermal conductivities the sample can have portions at widely different temperatures, i.e. with widely different magnetic susceptibilities. Kurti & Simon inferred that this was happening in the following experiment. Iron ammonium alum was cooled to a temperature below the maximum of magnetic susceptibility κ, so that with increasing temperature κ increases (Kurti, Lainé & Simon 1937). In this experiment, whereas a homogeneous heating by γ-rays duly produced an increase of κ, the temperature drift appeared as a decrease of κ. This is because the heat flow to the surface layers had warmed them above the inversion temperature, in fact to 1° where they contribute little to the average κ and the drift therefore appeared as a decrease of κ because only the surface layers are affected.

This is a particularly striking instance, but it is evident that the surface heat flow will produce inhomogeneity in temperature whenever the heat conductivity is low. Most experiments at very low temperatures use samples of inorganic salts either alone or mixed with other solids, and their conductivity is bad below about 1° K. (The only exception is in the use of 'capsules' of mixed salt and liquid He II, in which the good conductivity of the He II keeps the temperature uniform, and then these remarks do not apply.) Due regard must be had to this in all quantitative experiments, particularly long ones (e.g. measurements of specific heat) where the effect accumulates; it is impossible to correct for the temperature inhomogeneity, it must be kept sufficiently small to be negligible. An individual experiment can be continued only until the inhomogeneity becomes appreciable; it is then necessary to make a new demagnetization, or else to heat the sample in an alternating magnetic field. This latter procedure is based on the fact (Cooke & Hull 1937; Shire & Barkla 1939) that the heating of a paramagnetic substance by such a field is more rapid the lower the temperature. It is apparent that the result is an equalizing of the temperature throughout the sample, though at the expense of any readings in the temperature range where the a.c. heating took place. Frequent applications of this method, or frequent demagnetizations, are obviously to be avoided.

There are thus many reasons for desiring the best possible insulation. To some extent this may in effect be achieved by using a larger sample, when the effect of the heat flow (assumed to be independent of the size of the sample or at least to depend on its surface area rather than its volume) will be proportionately less. This artifice, as mentioned below, is only of limited application. The present paper is concerned with the more direct problem, the identification and removal of the causes of poor insulation.

2. Previous work on insulation

In looking for the source of heat flow at these low temperatures, it is necessary to consider the possible mechanisms. They are, first, heat transport through any gas still present round the samples (remaining from the helium 'exchange gas' used to carry away the heat of magnetization), and second, heat transport by conduction at points of contact. It will be shown later that heat transfer ordinarily occurs through both of these processes.

Direct contact was the more obvious potential source of heating, and attention had already been paid to this. In the early experiments of Kurti & Simon (1935 a) the compact sample was simply placed inside a container, and thus it touched the metal walls which were surrounded by a liquid helium bath at about 1° K. This served for qualitative experiments or for simple quantitative surveys (Kurti & Simon 1935 b). The method is unsatisfactory for more accurate measurements, e.g. of the thermodynamic scale, and for experiments with diluted salts, which produce the lowest temperatures but which necessarily have a low heat capacity. One difficulty with this simple arrangement is that the measured susceptibility is liable to decrease quite erratically; any vibration of the apparatus alters the points of contact of sample and container, and so instead of smooth susceptibility changes, sudden decreases or changes in rate of decrease are possible.

In their later Paris experiments (experimental details unpublished) Kurti & Simon mounted the sample, ellipsoidal in shape, between distance pieces of perspex or hollow pyrex. The insulation was effectively still further improved by using large samples, major axis 9·6 cm., minor axis 3·2 cm. They also developed a useful procedure for reducing to a minimum the amount of 'exchange gas' remaining around the sample. The gas is pumped out, before demagnetization of course, while the helium bath is at 1·5° K. Then, just before demagnetizing, the temperature of the bath is lowered to 1° K; this is effectively a 'baking out' of the metal vessel containing the sample, at the high temperature of 1·5° K. The temperature reached on demagnetization will correspond to about 1·5° K as the temperature of magnetization. The exact temperature of magnetization is not known, so that this method cannot be used in entropy measurements where the initial temperature must be known; also, of course, a lower temperature is attainable by proceeding in a straightforward manner from 1° K. Nevertheless, this technique can be valuable where removal of the helium gas is of prime importance.

Other insulation arrangements have been described, e.g. paper distance pieces by Shire & Allen (1938); but no systematic figures have been published for the heat flow under different conditions.

3. The experiments

We have made systematic measurements of the heat flow with various types of insulation; we have also investigated the possibility of utilizing what are effectively

distance pieces cooled to the same low temperature, which we have called 'guard rings'. Small samples only were used, because the magnet available was not suitable for large pole separation.

Here we may explain the method we have used throughout for calculating the heat flow. The effect of a heat flow is a decrease in susceptibility, and from the measured decrease we have to deduce the energy input. In doing so it is necessary to bear in mind that whatever process is responsible for the heat flow, it takes place primarily to some point or points on the surface of the sample. Because the sample has a bad thermal conductivity, the heat flow will warm part or all of the outer layers to $1°$ K where the conductivity becomes good again, without affecting the middle of the sample. The method of evaluating the energy input is thus to deduce from the susceptibility change the amount of substance heated to $1°$ K, and then to obtain the corresponding energy from the known specific heat curve. At temperatures nearer to $1°$ K where the conductivity improves, it may be that the sample is heated homogeneously. But here the specific heat curve follows an inverse square law, and it happens in consequence* that the same energy input corresponds to the same change of κ either on the assumption of uniform heating or on the assumption of some fraction of the substance being heated to $1°$ K. In all cases we have therefore made the evaluations on the assumption of some fraction only being heated, and that to $1°$ K.

Experiments on single Samples

In our first experiments the sample was a compact cylinder 5 cm. long and $1·2$ cm. diameter compressed from powdered iron ammonium alum. The sample was varnished thinly, immediately after making, with bakelite resin dissolved in acetone. This layer of varnish hinders any evaporation of water of crystallization, which would impair the paramagnetic behaviour of the salt; also, helium gas is likely to be adsorbed less on the varnish. The sample rested loosely inside a metal container from which it was separated by glass distance pieces. The container was surrounded by liquid helium at about $1°$ K. The temperature of the sample was measured by its susceptibility, as shown by its effect on the mutual inductance of two coils surrounding it.† The heat flow to such a sample at temperatures between $0·1$ and $0·6°$ K is about 25 ergs/min. and roughly independent of the temperature. With a homogeneous heat flow of this magnitude the sample would warm from

---

* Consider unit mass of the sample. At temperatures well above the characteristic θ for the substance used, $\kappa = a/T$ and c, the specific heat, $= b/T^2$. Assuming uniform heating, energy input $\delta Q = b\,\delta T/T^2 = b\,\delta\kappa/a$. Assuming a fraction f raised from T to T_1 (κ for this fraction being changed to κ_1), then $\kappa - \delta\kappa = \kappa(1-f) + f\kappa_1$. Therefore $f = \delta\kappa/(\kappa - \kappa_1)$. Therefore

$$\delta Q = \frac{\delta\kappa}{\kappa - \kappa_1} \int_T^{T_1} c\,dT = \frac{\delta\kappa}{\kappa - \kappa_1} b\left(\frac{1}{T} - \frac{1}{T_1}\right) = \frac{b\,\delta\kappa}{a}.$$

† This experimental arrangement is illustrated in figure 1 of a previous publication (Cooke & Hull 1937), except that the helium container was then of glass. The metal container used here is shown in figure 2 of the present paper.

0·2 to 0·5° K in 40 hr. This degree of insulation is good enough for many thermal investigations on the properties of such salts, and most of the early work in this laboratory was carried out in this way.

We next suspended the sample between two taut fibres, one attached to each end. The fibres are attached to a light metal cage as shown in figure 1. The cage is made from german-silver tube, 0·1 mm. thick, and has close-fitting brass caps.
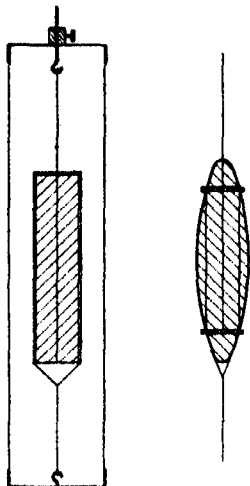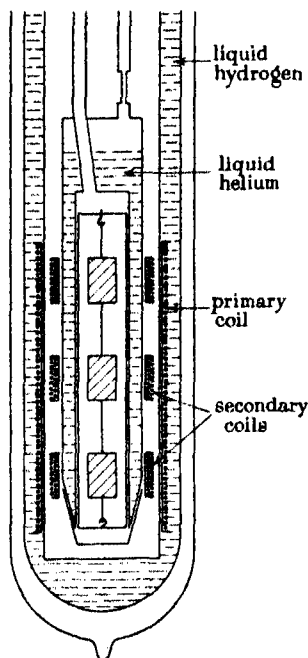
FIGURE 1   FIGURE 2

Most of the metal is cut away from the tube and the caps (*a*) to make the cage springy so that it will have good contact with the walls of the helium container and (*b*) to give low resistance to evacuation. The lower fibre is tied round the sample in the plane of the paper of figure 1, the upper fibre in the perpendicular plane; a tiny loop is left at the free end of each fibre. The method of assembly is then to lower the sample into the cage, fit the bottom loop on to the hook on the cap there, put the top cap on, fit the top fibre loop to the hook there and make the fibres taut by adjusting the top hook and its grub screw. The whole assembly can then be pushed into the inner space of the helium container, into which it just fits. Figure 1 also shows the method of suspending an ellipsoidal sample. Two thin perspex collars are used, the top one fastened to the bottom fibre and vice versa. With either sample it is advantageous to incorporate a light helical spring at the end of one of the fibres, particularly when the apparatus is subject to vibration.

We tried fibres 1·5 cm. long of (i) fused silica 0·2 mm. diameter, (ii) unspun artificial silk 0·35 mm. wide, and 0·08 mm. thick, and (iii) constantan 0·05 mm. diameter. In each experiment we used the 'baking out' procedure already described. With fused silica and artificial silk the heat flow was then 5 ergs/min., with constantan it was 10 ergs/min. In all the later experiments in this paper suspensions of artificial silk were used because of the comparative ease of handling them.

This simple type of suspension, a taut fibre at each end of the sample, the other ends of the fibres being at 1° K is now standard in this laboratory for ordinary experiments below 1° K. The insulation is sufficient for almost all ordinary measurements, and the use of very large samples is avoided.

Experiments using 'guard rings'

The next step was to find out to what extent this remaining heat flow of 5 ergs/min. was due to conduction along the fibres and to what extent it was due to heat transfer through the helium gas remaining in the container. For this purpose we
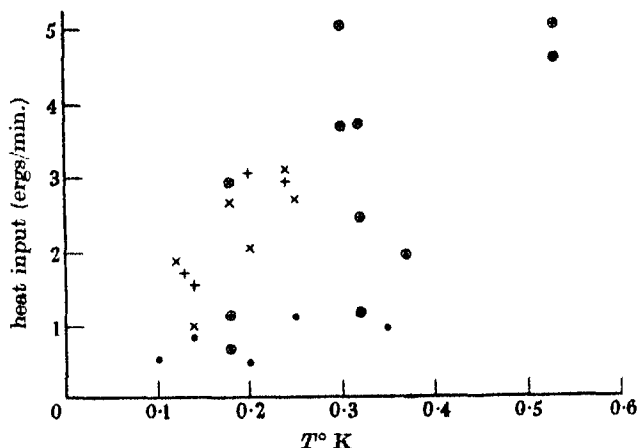
FIGURE 3

inserted half-way along each fibre a similar cylindrical sample of salt, also cooled to the very low temperature and thus acting as guard ring. The whole arrangement is shown in figure 2. The samples in these experiments were 3 cm. long and 1 cm. diameter. In order to measure the temperature of each sample it was necessary to have separate measuring coils for each. These were actually the secondary coils of the mutual inductances and there was only one primary coil common to all. We took the precaution of verifying that the resistances of these secondary coils stayed constant throughout an experiment (this is equivalent to saying that they were in good thermal contact with the helium container), so that the susceptibility of a sample could be deduced directly from the ballistic galvanometer deflexion.

The effect of any sample on the coils adjacent to its own, could be allowed for after a control experiment in which the middle sample was absent.

The results of the various experiments made with this arrangement are shown in table 1 and figure 3. Experiments were carried out with all three samples at the same temperature, with the middle one considerably warmer than the other two, in good vacuum obtained by 'baking out', and without pumping at all. It will be seen that under all these varying conditions the heat flow to the middle sample was about 1 erg/min. This considerable improvement in insulation means that in the simpler arrangement using a single sample, the heat flow of 5 ergs/min. was mainly due to heat transfer along the fibre. The possibility of its having been due to helium gas falling down the pumping tube (which was directly above) and giving up heat of condensation to the sample, is ruled out; because now, in conditions otherwise similar, the top guard ring receives only the same amount of heat as the bottom one.

TABLE 1

experiment	comment	sample		T° K	heat input ergs/min.
1	control experiment,	top,	iron amm. alum	0·25	2·7
	good vacuum	middle,	,,	0·25	1·0
2	,,	top,	,,	0·18	2·7
3	good vacuum	top,	,,	0·20	2·1
		middle,	,,	0·20	0·45
		bottom,	,,	0·20	0·8
4	,,	middle,	,,	0·14	0·8
		bottom,	,,	0·14	1·5
5	,,	top,	,,	0·115	1·85
		middle,	,,	0·105	0·5
		bottom,	,,	0·13	1·7
6	exchange gas	top,	,,	0·53	9·3
	7×10^{-4} mm.	middle,	,,	0·53	4·7
		bottom,	,,	0·53	5·1
7	,,	top,	,,	0·32	3·7
		middle,	,,	0·32	1·1
		bottom,	,,	0·32	2·5
8	,,	top,	,,	0·18	3·0
		middle,	,,	0·18	0·65
		bottom,	,,	0·18	1·1
9	exchange gas	top,	,,	0·295	5·1
	5×10^{-4} mm.	middle,	mn. amm. sulphate	0·37	1·9
		bottom,	iron amm. alum	0·295	3·7
10	good vacuum	top,	,,	0·24	3·1
		middle,	mn. amm. sulphate	0·35	0·9
		bottom,	iron amm. alum	0·24	3·0

Heat transfer along the fibre may be due entirely to ordinary conduction through the material of the fibre or there may be some contribution from condensed 'exchange gas' due to the 'creep' of He II along the surface of the fibre. Experiments

on heat transfer to vessels containing He II (Rollin & Simon 1939; Blaisse, Cooke & Hull 1939; Daunt & Mendelssohn 1939) have shown anomalous behaviour apparently due to heat flow but eventually traced to actual transfer of helium by the creep of a surface film. Such transfer might take place along the fibre if there were sufficient helium on its surface. But first we must see if there is any unexplained heating which would have to be attributed to this process. If we take the 5 ergs/min. obtained with the simple suspension to be due to ordinary conduction alone, we derive as mean values of the conductivity between 0·2 and 1° K, 10^{-5} cal./sec./cm. for silica or silk, and $2·5 \times 10^{-4}$ cal./sec./cm. for constantan. These values are not unreasonable in the light of measurements at somewhat higher temperatures (10^{-3} cal./sec./cm. for german-silver at 1° K), which makes it seem probable that 'creep' produces no great part of the heat transfer with the simple suspension.* Further experiments described below indicate that only with considerable helium gas present and able to condense on the fibre and sample is there any evidence of heat transfer that might be due to creep.

For the experiments in which the middle sample was warmer than the guard rings, it was made of manganese ammonium sulphate, while the guard rings were of iron alum. These two substances have different 'characteristic temperatures', and the table shows the temperatures reached by demagnetizing all three samples from the same initial field. It will be seen that the middle sample still warms at the rate of 1 erg/min. This shows that the 1 erg/min. is not being transmitted along the fibres, which are connected to the cooler guard rings. Heat of condensation of helium gas would account for it if helium gas were present continually at a pressure of 10^{-8} mm. of mercury. Such condensation would continually diminish the amount of helium gas present, and it might be expected that the heating would therefore decrease with time. In fact the heat flow remained constant for as long as it was measured, usually 2–3 hr. But this is not surprising when it is realized that the total quantity of helium gas that would be involved amounts to less than that in a monatomic layer on the walls of the container. There is certainly much more than this present inside the calorimeter, adsorbed on its walls, and remaining from the exchange gas used in cooling the samples from room temperature. In fact, surprisingly large quantities of helium have to be introduced into the calorimeter in order to have available the necessary pressure of exchange gas at 1° K (see, for instance, Giauque & MacDougall 1935; and Stout & Giauque 1938). Much of this gas is not removed by pumping so long as the calorimeter and its contents remain at 1° K or lower; it remains adsorbed on the surfaces with an equilibrium pressure much less than the ordinary saturated vapour pressure. We suppose then that at our very low temperatures there is on the samples a layer of helium with a negligible adsorption pressure, but that the finite equilibrium pressure of the helium

---

* There is the further possibility that some of the heat transfer might be due to ordinary conduction through adsorbed helium on the fibre, i.e. without the bulk transfer of helium associated with 'creep'. Our experiments give no evidence on the presence or absence of this effect.

on the calorimeter wall at 1° K leads to a continual transfer of heat to the samples. This explanation would account also for the fact that the heat flow to the middle sample is in general independent of its temperature.

It will be seen from figure 3 that the heat flow to the middle sample is not influenced by the presence of considerable helium gas at the moment of demagnetization (except for experiment 6, discussed below). In these experiments, 6 to 9, the gas pressure shown in the table was maintained during magnetization and not pumped at all. But actually the situation is not much altered; for now, on demagnetization, this gas pressure will be reduced by condensation on to the cold sample exactly as previously it was reduced by the combined effect first of pumping and then of condensation. The sample will not reach quite the same low temperature because of the latent heat of condensation, but its subsequent behaviour is not affected. In experiment 6, however, the presence of the same amount of exchange gas as in experiments 7 and 8 has exerted a profound effect on the heat flow to the middle sample. The essential point about this experiment is that all three samples are at relatively much higher temperatures; the situation is the same as if the samples in experiment 7 had all been heated to these higher temperatures. The ultimate effect of raising the temperature of the samples would be to restore the full exchange gas pressure. This moderate increase in temperature above that of experiment 7 does not restore the full pressure, but the equilibrium pressure of the helium adsorbed on the surface of the sample is higher than previously (the saturation vapour pressure at 0·55° K is 10^{-4} mm.). This higher pressure now contributes to the heat transfer between the calorimeter wall and the sample. It has always been realized that it is more difficult to preserve such intermediate temperatures than the very low ones; it adds considerably, for instance, to the skill required in establishing the absolute scale of temperature between 1° K and the lowest temperatures available. The 'baking out' process for removing helium gas is of great help in experiments involving temperatures between 0·4 and 1° K, particularly in specific heat measurements where heating has to be continued to high temperatures approaching 1° and where a knowledge of the initial entropy is not necessary.

The behaviour of the guard rings was not so uniform as that of the middle sample. Nevertheless, there are certain regularities to be noticed in figure 3. In each experiment with deliberately bad vacuum, the top sample warms faster than the bottom one. Initially the top sample warmed still faster for a few minutes because the exchange gas was not pumped at all, and so the gas in the pumping tube was falling down and condensing on this top sample. This condensation takes only a few minutes, and some other effect must be responsible for the fact that throughout the remainder of the experiment the top sample warmed faster than the bottom one. It may be that the condensation of such a large quantity of helium produces a film of He II sufficient to bring about this bigger heat flow. A He II film on a fibre with its ends at 1° K and say 0·3° K would show appreciable heat transport, even though the effective conductivity of the He II is no longer abnormally high

at 0·3° K (Kurti & Simon 1938). For the effect of a conductivity that varies with temperature is to make the temperature gradient non-uniform: most of the fibre would be at 1° K, and only a short length near the sample would lie in the range 0·3–1° K. It should be noted that a He II film on the fibre between guard ring and middle sample would have little effect because both ends of the fibre are below the temperature where the film has a high conductivity.

Except for this consistently greater heat flow to the top sample compared with that to the bottom one, the presence of gas did not have any marked effect on the 'guard rings' at temperatures below 0·3° K, as already noticed also for the middle sample. This is of considerable practical importance; for it means that once the bulk of the helium gas has been removed from the pumping tube to prevent this condensation on to the top sample just mentioned, further extensive pumping (e.g. the 'baking out' previously described) is of no advantage for experiments at the lowest temperatures, even when good insulation is desired. This would be true also for single suspended samples. The essential factor is the use of a suspension for good insulation, and the use of guard rings for still better insulation. On the other hand, in experiments at 0·4° K or higher, or in experiments with samples that rest against the container without a suspension, thorough pumping is certainly advantageous.

4. CONCLUSION

The experiments that have been described show that a heat flow of only 1 erg/min. is attainable and reproducible, and that this heat flow is then probably due to condensation on the sample of helium gas which leaves the wall of the calorimeter. It was at this stage that the experiments were interrupted in 1939. The sudden interruption accounts for the fact that some of them are not as conclusive as could be wished; nevertheless, we think it valuable to have a report on what has so far been achieved. When occasion permits, other useful experiments could be made (1) with the liquid helium container at different temperatures, thus varying the quantity of helium gas leaving the wall surrounding the samples, (2) by demagnetizing from higher temperatures where hydrogen gas could be used to remove the heat of magnetization, so that after demagnetization one would have the benefit of the much lower vapour pressure of hydrogen, and there could be no possibility of transfer of heat by a film of He II, (3) in which the middle sample is completely surrounded by a shield also cooled to the same very low temperature. The presence of such a shield, of metal or of paramagnetic salt, would introduce serious, but not insuperable, difficulties in the measurement of the temperature of the sample.

In the meantime it must be emphasized that the degree of insulation obtained here is considerable; whereas in the first recorded experiments on demagnetization the low temperature was preserved for only some minutes, this heat flow of 1 erg/min. would require more than 2 weeks to heat the sample from 0·2 to 0·5° K.

We wish to express our thanks to the Government Grant Committee of the Royal Society for aid in the purchase of apparatus. We are indebted to Professor F. Simon for many valuable discussions, and to Dr B. Bleaney and Dr N. Kurti for their generous help with the experiments.

REFERENCES

Blaisse, B. S., Cooke, A. H. & Hull, R. A. 1939 *Physica*, 6, 231.
Daunt, J. G. & Mendelssohn, K. 1939 *Proc. Roy. Soc.* A, 170, 423.
Cooke, A. H. & Hull, R. A. 1937 *Proc. Roy. Soc.* A, 162, 404.
Giauque, W. F. & McDougall, D. P. 1935 *J. Amer. Chem. Soc.* 57, 1175.
Kurti, N., Lainé, P. & Simon, F. 1937 *C.R. Acad. Sci., Paris*, 204, 675.
Kurti, N. & Simon, F. 1935 a *Proc. Roy. Soc.* A, 149, 152.
Kurti, N. & Simon, F. 1935 b *Proc. Roy. Soc.* A, 151, 610.
Kurti, N. & Simon, F. 1938 *Nature, Lond.*, 142, 207.
Rollin, B. V. & Simon, F. 1939 *Physica*, 6, 219.
Shire, E. S. & Allen, J. F. 1938 *Proc. Camb. Phil. Soc.* 34, 301.
Shire, E. S. & Barkla, H. M. 1939 *Proc. Camb. Phil. Soc.* 35, 327.
Simon 1939 *Conference on magnetism*. Strasbourg.
Stout, J. W. & Giauque, W. F. 1938 *J. Amer. Chem. Soc.* 60, 393.

The thermal diffusion of radon gas mixtures

By G. E. Harrison

Physics Department, University of Birmingham

In continuation of earlier experiments (Harrison 1937) in which the thermal diffusion in radon-hydrogen and radon-helium mixtures was measured, the thermal diffusion of mixtures of radon-neon and radon-argon has now been studied. The mean value obtained for the ratio of the proportion by volume of radon on the cold side at $0°$ C to that on the hot side at $100°$ C, after thermal diffusion, was 1.074 for radon-neon mixtures, and 1.008 for radon-argon mixtures.

In order to calculate the repulsive force field, F_{12}, between these two pairs of molecules, the present results were combined with measurements of ordinary diffusion of radon into neon and radon into argon (Hirst & Harrison 1939), and viscosity determinations at various temperatures of neon and argon (Trautz & Binkele 1930). The special theory, due to Chapman (1929), of thermal diffusion of a rare constituent in a binary mixture was used to derive F_{12}.

The values obtained for the repulsive force field between the dissimilar molecules at collision were:

$$F_{12} \text{ (radon-neon)} = 1.9 \times 10^{-51} d^{-6.1} = (d/d_0)^{-6.1}, \quad d_0 = 4.8 \times 10^{-9},$$

$$F_{12} \text{ (radon-argon)} = 2.1 \times 10^{-43} d^{-5.1} = (d/d_0)^{-5.1}, \quad d_0 = 4.3 \times 10^{-9},$$

d being the distance between the point centres of repulsive force and d_0 the value of d at which F_{12} is 1 dyne.

A comparison of the values obtained for the repulsive force index for radon-neon and radon-argon molecules with those obtained by Atkins, Bastick & Ibbs (1939) for binary mixtures of the first five inert gases shows that radon is the 'softest' of the inert gas molecules. Radon-argon molecules are the closest approach to the Maxwellian case yet studied experimentally.

INTRODUCTION

The repulsive force, F_{12}, between two dissimilar molecules may be represented by the equation

$$F_{12} = K_{12} d^{-s_{12}},$$

where s_{12} is the repulsive force index between the dissimilar molecules, d is the distance between the molecules which are regarded as point centres of repulsive force, and K_{12} is the repulsive force constant between these molecules.

In a special theory of thermal diffusion, Chapman (1929) has shown that when one component of a binary gas mixture is rare, F_{12} may be determined from thermal diffusion measurements for the mixture and the coefficient of viscosity of the lighter gas at various temperatures, together with the coefficient of diffusion of the two gases.

An account of an experimental study of thermal diffusion of radon-hydrogen and radon-helium mixtures in which radon was the rare and heavy constituent has already been given (Harrison 1937). The variation of the viscosity of hydrogen and helium over a wide range of temperature was known, but no experimental values of the coefficient of diffusion of radon into hydrogen or helium were available. Approximate values of these coefficients were therefore calculated from a formula given by Chapman (1918). Owing to uncertainty in the accuracy of these calculated values, only s_{12} was evaluated for radon-hydrogen and radon-helium molecules, K_{12} being much more sensitive than s_{12} to an error in the coefficient of diffusion. Experimental values for the coefficient of diffusion of radon into hydrogen, helium, neon and argon were later obtained by Hirst & Harrison (1939), who were thus able to derive the repulsive force between radon-hydrogen and radon-helium molecules.

Experiments on the thermal diffusion of radon-neon and radon-argon mixtures have now been completed. These experimental results have been combined with the diffusion measurements for these pairs of gases, and with the viscosity at various temperatures for neon and for argon (Trautz & Binkele 1930), to give the repulsive force between radon-neon and radon-argon molecules. Confirmation is thus obtained of the extremely 'soft' nature of the radon molecule, and it is shown that radon-argon molecules provide a close approach to Maxwellian molecules, for which there would be no thermal separation.

These measurements give information on the molecular field of radon which is particularly valuable, as it cannot be obtained directly in the usual way from viscosity measurements or an equation of state owing to the extremely small amount of gas available.

MODIFICATION OF EXPERIMENTAL METHOD

In the earlier experiments (Harrison 1937), the radon mixtures were prepared by breaking a capillary tube containing a known amount of radon in the diffusion apparatus, and allowing about 15 hr. for the radon gas mixture to become uniform by the process of ordinary diffusion. It has been shown (Hirst & Harrison 1939) that

if the adhesion of radon to the walls of a glass apparatus is to be avoided under the conditions of these experiments, the period of contact of the radon with the glass must not be greater than 7 hr. In the case of hydrogen-radon and helium-radon mixtures, the thermal separations were relatively large, so that a small amount of adhesion of radon in the thermal diffusion apparatus had little effect on the experimental results. The small thermal separations which were to be expected in the present experiments, however, especially when using radon-argon mixtures, made it necessary to avoid the long period of contact of radon with the walls of the diffusion apparatus. A capillary tube containing radon was therefore broken in neon or argon at atmospheric pressure, in an auxiliary apparatus, and left overnight for the radon gas mixture to become uniform. This mixture was put into the diffusion apparatus just before an experiment began. It will be shown that the necessary thermal diffusion data can be obtained without knowing the amount of radon put into the diffusion apparatus.

The principle of the method of measuring the thermal separations was similar to that already described for radon-hydrogen and radon-helium mixtures (Harrison 1937). As in these earlier experiments, the thermal diffusion apparatus consisted of a uniform, closed, glass tube 30 cm. long and of 1 cm. bore, divided into two parts by a wide-bore tap. The two parts formed the hot and cold sides of the apparatus. These two sides were also connected by means of a narrow capillary tube provided with a tap.

When the uniform gas mixture had been put into the diffusion apparatus, both taps were closed and one side was heated to 100° C and the other cooled to 0° C. This temperature difference was kept constant throughout the whole experiment. The tap in the capillary tube was opened for about 1 min. so that the pressures on the two sides could be equalized without any appreciable thermal separation taking place.

The gas apparatus was mounted throughout the experiment in front of a lead block provided with two apertures opposite the hot and cold sides, the one opposite the hot side being adjustable in width. The difference between the amounts of radon opposite the two apertures was found by observing the difference in the intensity of the γ-radiation passing through the apertures, after the radon had come into equilibrium with the products (Ra B and Ra C), which produce γ-rays. On the macroscopic scale of the present experiments, the solid products of radioactive change which were deposited on the walls of the diffusion apparatus were considered to be uniformly distributed. The difference in the γ-ray intensities from the two apertures was measured when equilibrium was established, i.e. 4 hr. after any change in the distribution of radon in the apparatus.

The two apertures were of equal width at the beginning of each experiment. After the pressure equalization but before thermal diffusion, the difference in the number of millicuries of radon in front of the equal apertures, n, was measured. This measurement was carried out by comparing the difference in the γ-ray activity from the apertures with the γ-ray activity of a standard radium source. The radium source

used was in the form of platinum needles and these were added to the hot side opposite the aperture in the lead block, so that the γ-ray intensity from the two apertures was approximately equal. The residual small difference in the γ-ray intensities from the apertures was then determined by observing the rate of drift of the electrometer system. The value of n is given by

$$n = a_2 v_2 - a_1 v_2,$$

where v_2 is the effective volume of radon opposite each aperture, and a_1 and a_2 are the number of millicuries of radon per c.c. on the hot and cold sides, after the pressure equalization. If T_1 and T_2 are the absolute temperatures of the hot and cold sides respectively, then

$$\frac{a_1}{a_2} = \frac{T_2}{T_1},$$

so that

$$a_2 v_2 = \frac{n}{1 - T_2/T_1}.$$

The number of millicuries of radon opposite the aperture on the cold side, $a_2 v_2$, was thus obtained.

After n had been measured, the radium source added to the hot side was removed, and the adjustable aperture was set at such a width that the γ-ray intensities from the two apertures were equal. The wide-bore tap was now opened for 1 hr. so that thermal separation could take place. By measuring the difference q in the amounts of radon in front of the apertures after thermal diffusion, a direct determination was made of the thermal separation. In this way, the duration of contact of radon with the walls of the diffusion apparatus was reduced to about 7 hr.

CALCULATION OF RESULTS

Using the same nomenclature as in the original account (Harrison 1937, p. 85), let x be the number of millicuries of radon per c.c. which thermally diffuse from the hot side of volume V_1, to the cold side of volume V_2. If v_1 is the effective volume of the aperture on the hot side after the adjustment in width of this aperture, then

$$q = x\left(v_1 + v_2 \frac{V_1}{V_2}\right).$$

The ratio of the proportion by volume of radon on the cold side to that on the hot side after thermal diffusion is given by

$$r = \frac{1 + \dfrac{x}{a_2}\dfrac{V_1}{V_2}}{1 - \dfrac{x}{a_1}}. \tag{1}$$

Eliminating x,

$$r = \cfrac{1 + \cfrac{V_1}{a_2 V_2}\left(\cfrac{q}{v_1 + v_2\cfrac{V_1}{V_2}}\right)}{1 - \cfrac{1}{a_1}\left(\cfrac{q}{v_1 + v_2\cfrac{V_1}{V_2}}\right)}.$$

As the effective amounts of radon opposite the two apertures were balanced after the pressure equalization by widening the aperture on the hot side,

$$a_1 v_1 = a_2 v_2, \quad \text{where} \quad a_1/a_2 = T_2/T_1.$$

Hence

$$r = \cfrac{a_2 v_2 + \cfrac{q}{1 + \cfrac{T_1 V_2}{T_2 V_1}}}{a_2 v_2 - \cfrac{q}{1 + \cfrac{T_2 V_1}{T_1 V_2}}}. \tag{2}$$

In the earlier experiments (Harrison 1937), the original proportion by volume c_1 of radon in the diffusion apparatus was known, and q, v_1, v_2 and $a_2 v_2$ were measured. The values of a_1 and a_2 could therefore be calculated and r was derived from formula (1). In the present experiments q and $a_2 v_2$ were obtained in the course of each experiment and r was calculated from formula (2). The necessary thermal diffusion data were thus obtained without knowing the amount of radon put into the diffusion apparatus.

RESULTS

In all the experiments, the original pressure of the mixtures was 1 atm., $T_1 = 100°$ C, $T_2 = 0°$ C and $V_1/V_2 = 1.16$. The results are shown in table 1, in which c_1 is the approximate relative proportion by volume of radon in the original gas mixture. This is only given to show that c_1 was very small in these experiments.

TABLE 1. EXPERIMENTAL RESULTS

gas mixture	c_1	$a_2 v_2$ millicuries of radon	q millicuries of radon	r^*
radon-neon	1.3×10^{-6}	15.4	1.03	1.068
	2.0×10^{-6}	23.7	1.81	1.078
radon-argon	1.4×10^{-6}	16.5	0.131	1.008
	1.8×10^{-6}	21.4	0.166	1.008
	2.3×10^{-6}	28.1	0.294	1.010
	1.5×10^{-6}	18.5	0.120	1.006

* As the value of r is almost unity, errors in r are less than one-tenth of the errors in q or in $a_2 v_2$.

The values of r obtained for radon-neon and radon-argon mixtures are smaller than the value obtained in the case of helium-radon mixtures (1·221) (Harrison 1937). There is a progressive decrease in r as a heavy molecule is substituted for a lighter one in a radon mixture. Since r increases with the total percentage thermal separation, it follows that this result is in agreement with that obtained in this laboratory by Atkins, Bastick & Ibbs (1939) using binary mixtures of the first five inert gases. This general agreement is striking, since the results for radon mixtures were derived by an entirely different experimental method.

It will be seen that r is nearly unity for radon-argon mixtures, which means that the thermal separation is extremely small, and radon-argon molecules are therefore a close approach to Maxwellian molecules for which r would be unity. Indeed, radon-argon molecules are the closest approach to the Maxwellian case yet studied experimentally.

Calculation of the field of force between radon-neon and radon-argon molecules

Chapman (1929) has shown that if c_1 is the proportion by volume of the rare and heavy constituent of a gas mixture, and c_2 is the proportion by volume of the second constituent such that $c_1 + c_2 = 1$, the equation which determines the steady distribution when the absolute temperature T varies in the direction of x is approximately

$$\frac{1}{c_1}\frac{\partial c_1}{\partial x} = -\frac{\alpha}{T}\frac{\partial T}{\partial x},$$

where α is a function of T, s_{12}, and s_2, the repulsive force index between the lighter, plentiful molecules.

If M_1 is the molecular weight of the rare component of the mixture and M_2 the molecular weight of the other component, then if M_2/M_1 and c_1 are both very small, a relatively simple expression may be derived for the first approximation to α, $\alpha(1)$. This expression is given in the earlier paper (Harrison 1937).

In the present gas mixtures, however, M_2/M_1 is not small, so that a more general expression has to be used for $\alpha(1)$. In one or two special cases only can this expression be reduced to a manageable form. One of these cases is when $c_1 \to 0$. In this case, the equation for $\alpha(1)$ when the molecules are supposed to possess spherical symmetry is

$$\alpha(1)_{c_1\to 0} = \tfrac{5}{2}\mu_{12}M_1^{-\frac{1}{2}}\alpha_{01}\left[\frac{\mu_{12}\alpha_{11} + \mu_{21}^{\frac{1}{2}}\alpha_{1-1} - \mu_{21}\alpha_{-1-1'}}{\alpha_{11}\alpha_{-1-1'}}\right], \qquad (3)$$

the nomenclature being the same as that used by Chapman (1929). The values of μ depend only on the mass ratio between the molecules, while those of α may be calculated from formulae given by Chapman (1929) if s_2 is known.

It is also shown by Chapman (1929) that

$$\log_e r = \frac{\alpha}{\beta}\left\{\left(\frac{T_1}{T_2}\right)^\beta - 1\right\}, \qquad (4)$$

where β is a constant for a given pair of molecules, which is given by

$$\beta = \frac{1}{2}\left\{\frac{s_{12}-5}{s_{12}-1} - \frac{s_2-5}{s_2-1}\right\}. \tag{5}$$

If the molecular weights of both gases are known, together with the coefficient of diffusion of these gases, and the coefficient of viscosity of the plentiful constituent at various temperatures, equations (3), (4) and (5) may be combined to give the repulsive force index, s_{12}.

When s_{12} has been calculated, the repulsive force constant between the dissimilar molecules, K_{12}, may also be evaluated. The theoretical formulae necessary for this calculation are summarized in the paper by Hirst & Harrison (1939). Thus, the repulsive force between the dissimilar molecules, F_{12}, may be obtained.

For the purpose of this calculation, the following data were used.

(1) *Neon-radon mixtures*

The coefficient of viscosity of neon was taken as 308×10^{-6} c.g.s. unit at 15° C (Trautz & Binkele 1930), and the repulsive force index as 13·5. The coefficient of diffusion of radon into neon as determined by Hirst & Harrison (1939) was 0·217 cm.² sec.⁻¹ at 15° C and 76 cm. mercury. The mean value of r for radon-neon mixtures from the present experiments is 1·074, while $M_1 = 20\cdot2$ and $M_2 = 222$.

(2) *Argon-radon mixtures*

The coefficient of viscosity of argon was assumed to be 218×10^{-6} c.g.s. unit at 15° C (Trautz & Binkele 1930), and the repulsive force index $s_2 = 7\cdot4$. The value 0·092 cm.² sec.⁻¹ at 15° C and 76 cm. mercury was used for the coefficient of diffusion of radon into argon (Hirst & Harrison 1939), and the mean value of r for radon-argon mixtures obtained from the present measurements is 1·008. $M_1 = 39\cdot8$.

The following values were obtained for F_{12}:

$$F_{12}\text{(neon-radon)} = 1\cdot9 \times 10^{-51} d^{-6\cdot1} = (d/d_0)^{-6\cdot1}, \quad d_0 = 4\cdot8 \times 10^{-9};$$

$$F_{12}\text{(argon-radon)} = 2\cdot1 \times 10^{-43} d^{-5\cdot1} = (d/d_0)^{-5\cdot1}, \quad d_0 = 4\cdot3 \times 10^{-9};$$

d being the distance between the point centres of repulsive force and d_0 the value of d at which F_{12} is 1 dyne.

DISCUSSION

Atkins *et al.* (1939) have given an account of the measurement of thermal diffusion in binary mixtures of helium, neon, argon, krypton and xenon. By application of the theory of thermal diffusion developed by Chapman (1918), in which neither component of the mixture is rare, the repulsive force index s_{12} has been evaluated for each pair of molecules. An abstract of the values obtained for s_{12} and the corresponding values obtained for radon mixtures is given in table 2.

It will be seen from the table that the effect of substituting a heavier monatomic molecule in any binary combination is to reduce the index of repulsive force, as pointed out by Atkins *et al.* (1939). The results of the present experiments confirm

G. E. Harrison

that radon fits into this scheme by giving the lowest value of s_{12} in each of the three series.

TABLE 2. s_{12} FOR BINARY MIXTURES OF THE INERT GASES
AT ORDINARY TEMPERATURES

gas mixture	s_{12}	gas mixture	s_{12}	gas mixture	s_{12}
He–Ne	11·4	—	—	—	—
He–A	9·0	Ne–A	7·9	—	—
He–Kr	8·7	Ne–Kr	7·6	A–Kr	5·7
He–X	8·3	Ne–X	7·0	A–X	5·6
He–Rn	6·5	Ne–Rn	6·1	A–Rn	5·1

A comparison of the values of s_1 or s_2 deduced from viscosity measurements on simple gases with the values of s_{12} obtained from thermal diffusion experiments, in which neither component is rare, shows that the values of s_{12} obtained from diffusion are generally lower than would be expected from viscosity measurements. Grew (1941) has shown that this disparity is due to an approximation which has to be used in order to derive s_{12} from the thermal diffusion measurements. Using an isotopic mixture of hydrogen and deuterium, Grew was able to avoid this approximation (because of the simpler force fields between the different molecules), and he then obtained agreement between the repulsive force index derived from thermal diffusion measurements and that obtained from viscosity. The values of s_{12} derived by Atkins *et al.* for binary mixtures of the inert gases given in table 2 are therefore probably too small, the error in the higher values of s_{12} being greater than in the smaller ones. The present results obtained for radon mixtures are independent of the assumption which is necessary in the general case of a mixture in which neither component is rare. Hence the values of s_{12} for radon mixtures still remain the smallest in each group and the general scheme shown in table 2 is not affected.

In view of the close approach of radon-argon molecules to the Maxwellian case ($s_{12} = 5$, $r = 1$), it would be of special interest to make thermal diffusion measurements for radon-krypton and radon-xenon mixtures. It is possible that s_{12} may be less than 5 for these pairs of molecules, and we should then expect the thermal separations to be reversed, the heavier gas tending to move to the hot side.

My thanks are due to Professor M. L. E. Oliphant, F.R.S., for the facilities which he has provided for carrying out this work. I am also indebted to Dr T. L. Ibbs for his valuable advice and criticism.

REFERENCES

Atkins, B. E., Bastick, R. E. & Ibbs, T. L. 1939 *Proc. Roy. Soc.* A, **172**, 142–158.
Chapman, S. 1918 *Phil. Trans.* A, **217**, 115–197.
Chapman, S. 1929 *Phil. Mag.* **7**, 1–16.
Grew, K. E. 1941 *Proc. Roy. Soc.* A, **178**, 390–399.
Harrison, G. E. 1937 *Proc. Roy. Soc.* A, **161**, 80–94.
Hirst, W. & Harrison, G. E. 1939 *Proc. Roy. Soc.* A, **169**, 573–586.
Trautz, M. & Binkele, R. E. 1930 *Ann. Phys., Lpz.,* (5), **5**, 561–580.

The structure of graphite

By H. Lipson and A. R. Stokes, *Cavendish Laboratory, Cambridge*

(Communicated by Sir Lawrence Bragg, F.R.S. -- Received 10 *April* 1942)

A new structure of carbon is proposed in order to account for the occurrence of extra lines on X-ray powder photographs of graphite. It has hexagonal layers similar to those of graphite, but arranged in a different sequence. About 14 % of the new structure is present in the samples examined.

INTRODUCTION

The structure of graphite is given in the literature as hexagonal with four atoms in the unit cell (Hassel & Mark 1924; Bernal 1924; Ott 1928); the dimensions of the unit cell are $a = 2\cdot456$ A, $c = 6\cdot696$ A (Trzebiatowski 1937). This structure, however, does not account for some faint lines that occur on X-ray powder photographs of well-crystallized graphite (Taylor & Laidler 1940). These lines have attracted attention because of their very general occurrence; they are given by graphites of many different origins, natural and artificial. For instance, they occur on a photograph that Rooksby (1940) gives as typical of natural graphite, and Taylor & Laidler (1940) found that they are given by graphite crystallized in an arc. Purification of the specimens does not remove the lines, but they do disappear after long periods of digestion with a mixture of concentrated sulphuric and nitric acids.

Finch & Wilman (1936) had observed similar extra lines on electron-diffraction photographs, and had accounted for them as secondary diffraction phenomena for plates only a small number (3, 4, or 5) of unit cells thick. Secondary maxima should appear at intervals that are submultiples of the intervals between the primary maxima, and so the extra lines were given simple fractional l indices. While this explanation may be plausible for electron-diffraction photographs, the sharpness of the lines on X-ray photographs shows that here we are dealing with much bigger crystals.

Taylor & Laidler (1940) suggested that the extra lines may be another manifestation of the 'diffuse' reflexion of X-rays, but Lonsdale, Knaggs & Smith (1940) showed that this cannot be so.

Edwards & Lipson (1942) thought that they may be due to defects in the graphite structure similar to those found in cobalt (Edwards, Lipson & Wilson 1941), but occurring at regular intervals. It was in order to test this suggestion that the present work was undertaken.

POSITIONS OF THE LINES

Measurements were made of a powder photograph, kindly supplied by Dr A. Taylor, of graphite crystallized by arcing. In figure 1 these measurements are

plotted against the angles calculated for the various indices shown, and it will be seen that the positions of the extra lines conform with those for simple fractional l indices. These indices are all multiples of $\frac{2}{3}$, and the theory that the lines are due to periodic repetition of faults would require that such faults should occur after every third unit cell. This theory, however, cannot account for the absence of odd multiples of $\frac{1}{3}$.
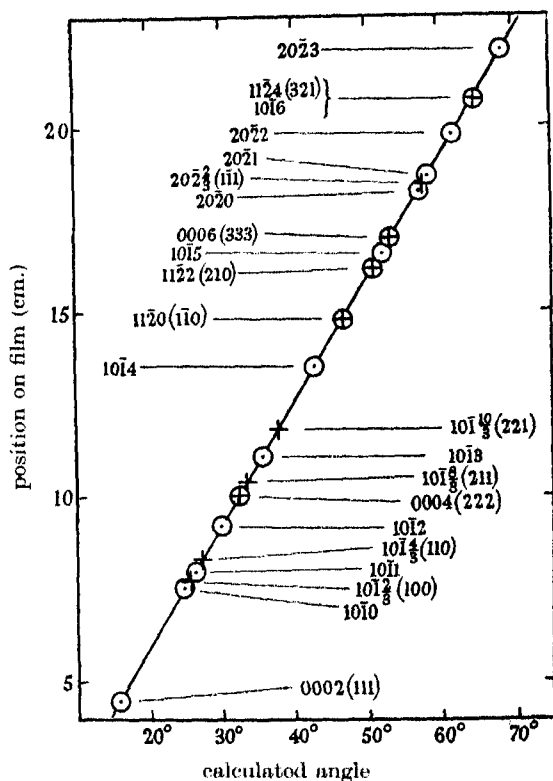

FIGURE 1. Positions of the lines. Rhombohedral indices (see text) are given in brackets.

PROPOSED STRUCTURE

A much simpler way of explaining l indices that are multiples of $\frac{2}{3}$ is to postulate a unit cell with a c-axis $\frac{3}{2}$ times as long as the usual one. The extra lines would then represent the effect of a small amount of such a structure mixed with the ordinary one.

Now the ordinary graphite structure consists of layers of carbon atoms forming hexagons with shared edges; each layer has half its atoms directly above atoms in the layer below, and the other half above the centres of hexagons. The complete structure is formed by repeating these layers in such a way that alternate ones are similar. In figure 2 two such layers are shown by thick lines and by double
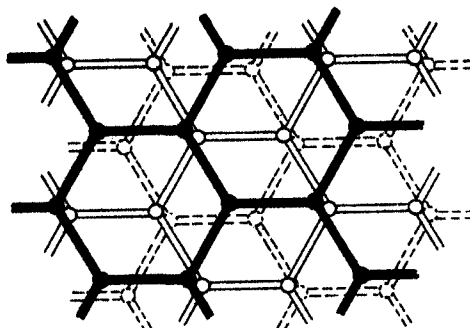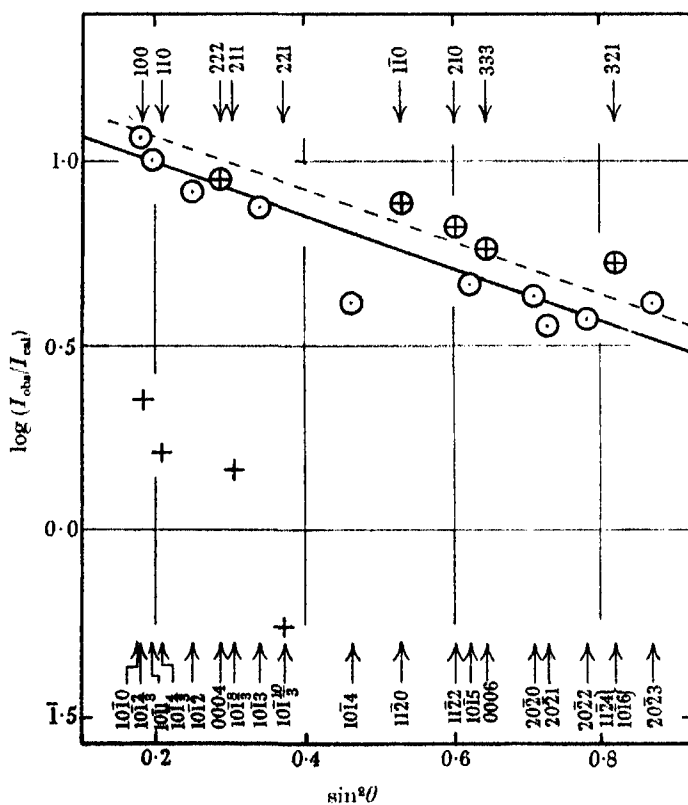
FIGURE 2. Arrangement of hexagonal layers (see text).

FIGURE 3. Photometry measurements. Hexagonal indices are shown at the bottom, and rhombohedral indices at the top, of the figure.

lines. It will be seen, however, that there is a third type of layer, shown by broken lines, which is symmetrically related to the other two. A structure may therefore be imagined in which these *three* layers repeat successively, and it is this structure that we propose as the origin of the extra diffraction lines.

It is obvious that the structure is very closely related to that of ordinary graphite. The strongest lines from it are coincident with the strongest graphite lines, and that is the reason why so few lines from it are visible. This gives a possibility of obtaining additional evidence for its existence, for measurements of intensities should show whether these strong lines—those with $(h-k)/3$ integral—are relatively stronger than the others.

The results of photometry of one film are shown in figure 3; in this figure a function of the intensities that should be linear in $\sin^2 \theta$ (Bradley & Lu 1937) is plotted against $\sin^2 \theta$. The lines from the ordinary structure are shown by circles, and those for the new structure by crosses, the lines common to the two being marked with the two signs superimposed. It will be seen that the intensities of these lines do lie above those of the others.

The agreement, however, is not quantitatively correct, as the increase in the intensities of the common lines is larger than it should be. This is shown by the broken line in figure 3, which is that upon which the intensities of the common lines should lie if the amount of structure present were that given by its separate lines; this point will be discussed later.

DESCRIPTION OF THE STRUCTURE

As we have stated, the structure may be referred to the hexagonal axes $a = 2.456$ A, $c = \frac{3}{2} \times 6.696$ A $= 10.044$ A. The atoms are in positions $0\ 0\ 0$, $\frac{1}{3}\ \frac{2}{3}\ 0$, $0\ 0\ \frac{1}{3}$, $\frac{2}{3}\ \frac{1}{3}\ \frac{1}{3}$, $\frac{1}{3}\ \frac{2}{3}\ \frac{2}{3}$, $\frac{2}{3}\ \frac{1}{3}\ \frac{2}{3}$. This structure, however, is really rhombohedral, with $a = 3.635$ A, $\alpha = 39.49°$. The atoms are at $\pm (u, u, u)$, with $u = \frac{1}{6}$, and the space group is $R\bar{3}m$. The parameter u cannot, of course, be determined exactly; we have chosen this value as it gives plane hexagonal rings. Actually the value $u = 0.164$ would give slightly better agreement with the intensities, and this would mean that the atoms were ± 0.03 A out of the planes. The intensities, however, are too small for this point to be stated definitely.

PROPORTIONS OF THE STRUCTURES

It was considered advisable to verify that the enhancement of the lines with $(h-k)/3$ integral did not occur for the lines from the ordinary structure, and so we attempted to remove the new structure by digestion with concentrated sulphuric and nitric acids. However, photometry of the resulting photograph, which showed only faint traces of the extra lines, still gave the same enhancement of the lines with $(h-k)/3$ integral. It thus appears that the action of the acids is to rearrange the layers of atoms, rather than to remove the new structure; this would leave the

lines common to the two structures unaffected, but might make the others so broad that they would be undetectable (Wilson 1942). Support to these ideas is given by the observation that the specimen did not lose weight during the transformation. Also lines with l non-zero became broadened, indicating that the regular arrangement of the planes is affected; other lines, especially $11\bar{2}0$, remain almost unaltered, showing that the planes themselves remain perfect.

The considerations suggest an explanation of the enhancement of the lines with $(h-k)/3$ integral, for there may always be a certain amount of the structure with disordered packing present. The data given in figure 3 could be explained completely by a specimen that contained 14 % of the new structure, 6 % of the disordered structure, and 80 % of the ordinary structure.

The general occurrence of the new structure is remarkable. It occurs in samples of natural graphite from Ceylon, Bavaria, and Travancore, and in graphite extracted from the 'kish' which occurs in the casting of carbon-rich steels; to visual examination of the photographs, the same proportions of it seem to be present as in the artificial sample used by Dr Taylor.

REFERENCES

Bernal, J. D. 1924 *Proc. Roy. Soc.* A, **106**, 749.
Bradley, A. J. & Lu, S. S. 1937 *Z. Kristallogr.* **96**, 20 (p. 24).
Edwards, O. S. & Lipson, H. 1942 *Proc. Roy. Soc.* A, **180**, 268.
Edwards, O. S., Lipson, H. & Wilson, A. J. C. 1941 *Nature, Lond.*, **148**, 165.
Finch, G. I. & Wilman, H. 1936 *Proc. Roy. Soc.* A, **155**, 345.
Hassel, O. & Mark, H. 1924 *Z. Phys.* **25**, 317.
Lonsdale, K., Knaggs, I. E. & Smith, H. 1940 *Nature, Lond.*, **146**, 332.
Ott, H. 1928 *Ann. Phys., Lpz.*, **85**, 81.
Rooksby, H. L. 1940 *J. R. Soc. Arts*, **88**, 308 (p. 330).
Taylor, A. & Laidler, D. 1940 *Nature, Lond.*, **146**, 130.
Trzebiatowski, W. 1937 *Roczn. Chem.* **17**, 73 (or *Chem. Abstr.* **31**, 4177).
Wilson, A. J. C. 1942 *Proc. Roy. Soc.* A, **180**, 277.

Some experiments on the trichromatic theory of vision

By H. V. WALTERS

Imperial College of Science and Technology, Technical Optics Section

A repetition of an investigation carried out by Wright in 1934 is made with an improved design of the Wright trichromatic colorimeter. The technique is described and the assumptions in the method for determining the fundamental response curves discussed. New observations are reported in which the generalized form of von Kries's law of coefficients is found to break down, and the departures from the coefficient law are analysed. Possible physiological causes for the breakdown are discussed. By a modification in the calculations, a new attempt has been made to locate the colours in the trichromatic colour chart corresponding to the three fundamental responses, and to derive the response curves.

Calculations were first made in the W. D. Wright trichromatic system and were transformed into the rectangular uniform chromaticity scale system of Breckenridge & Schaub, showing the caution required in interpreting the statistical spread of the results.

The ultimate shape of pebbles, natural and artificial

By Lord Rayleigh, F.R.S.

Flint pebbles that have been worn down enough to be symmetrical may be spherical, or may approximate to prolate or oblate spheroids or to ellipsoids. When, however, the section is at all elongated, it is not as a rule accurately elliptical, but except at the axial points it lies entirely outside an ellipse adjusted to the same axes. Thus, if one of the axes is much smaller than the other, the pebble is much flatter than an ellipsoid.

Considering the quasi-spheroidal pebbles, whether prolate or oblate, these are always flattened at the poles, and the very oblate ones become tabular or even concave at the poles. These statements hold for flint pebbles, but a large quartzite pebble is figured which is pretty accurately an oblate spheroid, the meridional sections being truly elliptical.

Experiments are described with chalk pebbles, initially shaped as prolate or oblate spheroids, and the change of figure under abrasion is observed. This depends in some degree on what is the abrasive. Steel nuts, nails ('tin tacks') and small shot were used. In general the axes tend to approach equality, but not rapidly enough for the spherical form to be attained before the pebble has disappeared. The form initially spheroidal becomes flattened at the poles just like the natural flint pebbles, and may become concave, as flints sometimes do.

The circumstances determining the rate of abrasion at any point are considered, and it is shown that this abrasion cannot be merely a function of the local specific curvature. The figure at points other than the one under consideration comes into question, so that the problem in this form has no determinate answer. It is shown by simple mechanical considerations how the concave form arises.

The large majority of pebbles found in nature though rounded in outline are of unsymmetrical shape. This is no doubt mainly due to their not having undergone enough attrition to attain symmetry, but also in some cases to the material not being homogeneous—for it is sometimes found in experiment that the original artificially imposed symmetry is lost in the course of attrition. The present paper only deals with that minority of natural pebbles which are symmetrical, and with the experimental study of attrition, starting with definite symmetrical shapes.

My own recent observations of natural pebbles have been chiefly made on the flint deposits classed as glacial drifts, the beaches not being now accessible. Spherical pebbles are occasionally met with, but they are so rare in my district (Terling, Essex, not far from Chelmsford) that they are sometimes put aside as minor curiosities. I have found or been given five or six of them in the course of a year's interest in the subject, in a district where the gravel is extensively put down on roads of occupation, farmyards and the like. Pebbles approximating to oblate spheroids are fairly common, especially in small sizes, 2 to 3 cm. in diameter. Prolate forms of revolution are very rare, indeed practically non-existent. Long finger-like natural shapes of flint are not uncommon in the chalk and in the drifts, and it might be imagined that they would yield approximate prolate spheroids by attrition, but in fact elongated pebbles when found are practically always wanting

in symmetry. It would seem that this shape is unstable, and loses symmetry instead of gaining it in course of attrition. Experimental observations casually made tend to confirm this view.

Coming now to a closer view of the exact shape, it is often found that the spheres are fairly accurate. I have seen a spherical quartzite pebble from Woolshed, Beachworth, Victoria, Australia that was almost as accurately shaped as a billiard ball. It would seem that if what may be loosely described as the three principal axes are not very different initially (cube, e.g.) then the spherical form is continually approached as attrition proceeds. Oblate forms very close to the sphere are certainly not often met with. I have, in fact, never found one with the polar axis as much as 0·7 of the equatorial diameter.

The principal sections of a pebble, if not circular, are seldom accurately elliptical. An example is seen in figure 1 (plate 1), which represents the section ac of a fairly symmetrical pebble having its axes $2a = 2\cdot95$ cm., $2b = 2\cdot12$ cm., $2c = 1\cdot49$ cm. The section was drawn by projecting the enlarged silhouette of the pebble optically, and going round the shadow with a pencil. An ellipse was drawn with the same axes $2a$ and $2c$. It will be seen that the ellipse lies within the pebble section, except at the four points on the axes, when it was adjusted to fit. The figure shown was reduced photographically from the enlarged graph.

This inscribed ellipse was drawn as the simplest way of showing that the section was not elliptical. The circumscribed ellipse which could only be drawn by tentative methods would show more directly how the actual pebble could be derived by local abrasion or figuring (or rather disfiguring) of an ellipsoid.

Pebbles like this one, having three unequal axes, exhibit a flattened shape in each of the principal sections, the curve bulging away at four places from the inscribed ellipse, and naturally this is more apparent the farther the section departs from a circular form. Whatever the significance of the flattened form, it is most simply exemplified when one section is circular, so that no unnecessary complications arise.

Such pebbles then are usually flattened at the poles* and in the case of flints they are invariably flattened, so far as my experience goes.

The following photographs are silhouettes (shadows cast by a distant arc lamp) in actual size, the axis of the figure being in each case vertical. Figure 2 (plate 1) is the nearest approach to a prolate figure of revolution that I have been able to find. The axial measurements are 3·18 cm., 2·21 cm., 1·71 cm. It is obvious by inspection that the ends are much blunted compared with the circumscribed ellipse.

Figure 3 (plate 1) is an oblate figure of revolution, again much flattened in the axial direction.

There exist, however, oblate pebbles, which are not flattened in this way, but approximate closely to oblate spheroids. Figure 4 (plate 1) shows a pebble of

* In popular language the earth is said to be flattened at the poles. This means flattened from the sphere to the oblate spheroid. The pebble is flattened beyond this, the section being no longer elliptical.

quartzite which I owe to the kindness of my friend, Sir D'Arcy Thompson, F.R.S. He was unable to say with certainty where he collected it, but thinks it most probable that it was on the banks of the River Spey in Inverness-shire. Figure 5 (plate 1) is an ellipse of nearly the same dimensions as the pebble section for comparison. In the same collection were two other similarly shaped pebbles of some other fine-grained igneous or metamorphic rock, not precisely identified.

Returning to the oblate flint pebbles, it is found that some of the thinnest ones are not merely flat at the poles, but actually concave. Figure 6 (plate 1) illustrates this. The concavity cannot be accurately photographed in profile, for obvious geometrical reasons, but a straight-edge has been placed over the concave surface, and with suitably adjusted positions of light and camera lens scattered light passes below the straight-edge, proving concavity. The other side is also concave though the illumination cannot well be arranged to show this in the same picture. The pebble measures $3.0 \times 2.2 \times 0.8$ cm., and the radii of concave curvature are estimated at 19 cm. and 30 cm.

In the case of natural pebbles there are no means of knowing definitely by what stages the individual pebble has attained its present shape. In particular we cannot say a priori whether a flattened surface has not been formed by the sides having been embedded, and the face alone exposed to attrition. It is true that in the case of a symmetrical pebble, we should have to assume that the stone had been turned over and the other side similarly flattened in its turn: but this cannot be clearly excluded.

Here comes in one great advantage of experimental conditions. They can be so arranged that no accident can have any preponderating influence and the influences at work are those which the statistical conditions themselves dictate. Further, the changing shapes of one particular pebble can be examined as attrition proceeds. Under natural conditions this cannot be done, and there is no definite information as to the shape which preceded that actually observed.

Many experiments have been made in the past on the formation of pebbles by abrasion, and some references are given at the end of this paper (Wentworth 1919, Showe 1932, Bullows 1939, Krumhein 1941). Some of the investigators are chiefly interested in the question of what distance the pebble must travel under water to produce a well-rounded form in the various kinds of material, and generally to see what can be inferred about the geological history of similar pebbles occurring in nature. I have not been able to find much about the questions here discussed, though it is of course widely recognized that the typical form of pebbles is flattened, and also that spheres are rare, and are not produced unless the original pieces have their three principal axes (using the terms in a loose general sense) nearly equal. It may be remarked that experiments on slate, or indeed any rock which has distinctive planes of cleavage or bedding, are not helpful in the present discussion.

The method of investigation used is to begin with pebbles of a definite geometrical form, free from sharp edges, and to observe the modification of shape as

attrition proceeds. There is little choice as to the initial form. An ellipsoid is clearly indicated. If we were to go beyond this we should have to consider surfaces of the fourth degree, which is scarcely practicable.

It is useful to consider in the first instance what relation must hold between the specific curvature at any point of an ellipsoid and the linear rate of attrition measured perpendicular to the surface in order that the shape should be unaltered by the process of attrition.

A general discussion of the curvature of surfaces is given, e.g. by Thomson & Tait (1879, p. 101). The reader may be reminded that at any point of a surface there are two principal curvatures. These are the curvatures of plane sections normal to the surface, these sections being chosen so as to have maximum and minimum curvature. It may be shown that they are mutually perpendicular. The specific curvature of the surface can be calculated as the product of these two principal curvatures.

In the particular case of the ends of the axes of an ellipsoid the specific curvatures are easily found. If a, b, c are the semi-axes, the specific curvatures are

$$a^2/b^2c^2, \quad b^2/a^2c^2, \quad c^2/a^2b^2.$$

Multiplying through by $a^2b^2c^2$, we see that these are in the ratio

$$a^4 : b^4 : c^4.$$

If, therefore, in any given case the ratios of the axes $a:b:c$ are found to be unaltered by abrasion, it follows that in this case the rate of abrasion at their ends must be as the fourth roots of the specific curvatures at these ends. It can be shown that the same holds at other points of the ellipsoid though naturally the geometry in this general case is not quite so simple. (See appendix at the end of the paper.)

The questions of special interest in connexion with the abrasion of ellipsoids appear to be all sufficiently exemplified by prolate and oblate spheroids, which are much easier to shape in the first instance and to examine and discuss afterwards. Accordingly these have been used exclusively in the experimental work.

Chalk was used as the material because it is soft. This makes it easy to shape the spheroids in the first place, and the desired amount of abrasion can be carried out in periods of at most a few hours, and often less. The method of producing chalk spheroids is simple. An elliptic aperture is made in a piece of sheet brass about 1 mm. in thickness. This would best be done with an oval chuck, as used in ornamental turning. In the absence of such a tool, the ellipse may be scribed on the brass, which can be done well enough for the purpose by the common method of a loop of (linen) thread round two drawing pins. These pass through holes in the brass into a board below. The aperture is then filed out up to this mark. Having prepared the elliptical template, a piece of chalk is sawn and rasped to near the desired shape, and then revolved in the fingers against the template used as a cutting tool until it will just go through. In a few minutes a very satisfying

spheroid can be produced. The same template can of course be used to make a prolate or oblate spheroid of a given ratio of axes. The chalk used was from the Cherry Hinton pit near Cambridge. It occasionally contains shells or other hard inclusions which resist abrasion more than the body of the material, and produce local irregularities, but the trouble from this cause is not serious. The abrasion was
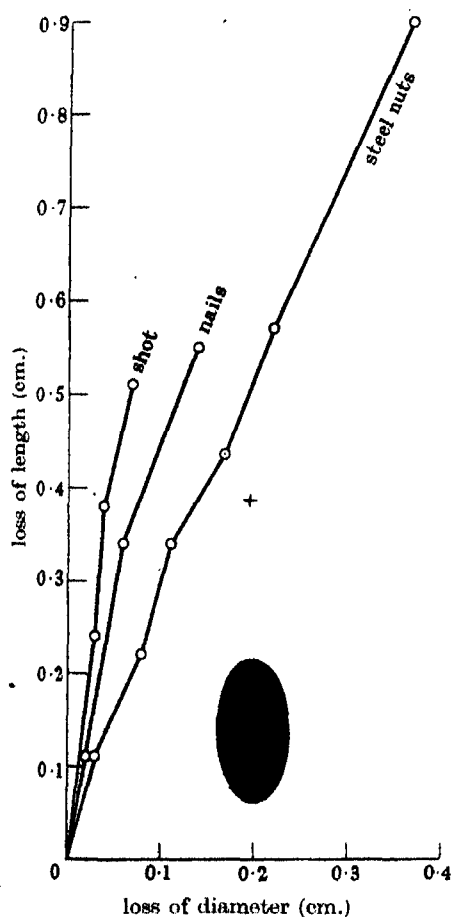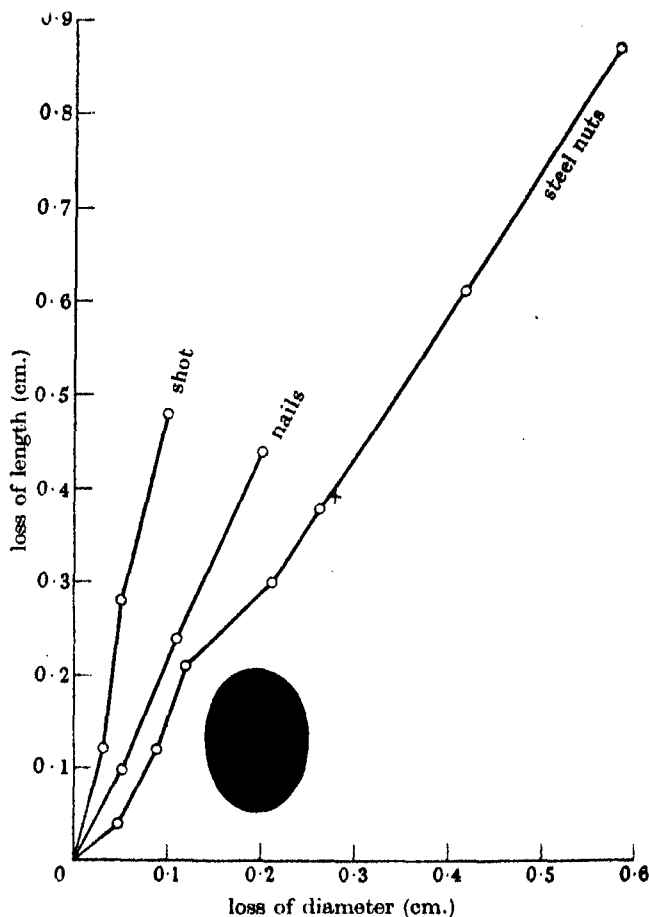
FIGURE 7. Prolate.   FIGURE 8. Prolate.

carried out in the absence of water. This is a departure from the conditions which generally prevail in the formation of natural flint pebbles, but it may be doubted whether using the chalk wet would really imitate these conditions more closely, for unlike flint, chalk gets soft and muddy under water. In fact, wet chalk is too soft for satisfactory measurement as attrition proceeds and to stop and dry it on each occasion would seriously delay the work. Further it is found that patches of local softness leading to irregular figure are more noticeable when the material is wet.

The shaped chalk pebble was placed in a metal box measuring $18 \times 10 \times 8$ cm. with either 470 g. of steel hexagon nuts somewhat various but weighing on the average 7·5 g. each; or 420 g. of small nails ('tin tacks'), length 10 mm., weight 0·3 g.; or 1200 g. of small shot, diameter of pellets 2·5 mm., weight 0·1 g. There was nothing special about these quantities, which were originally taken at random, and adhered to.
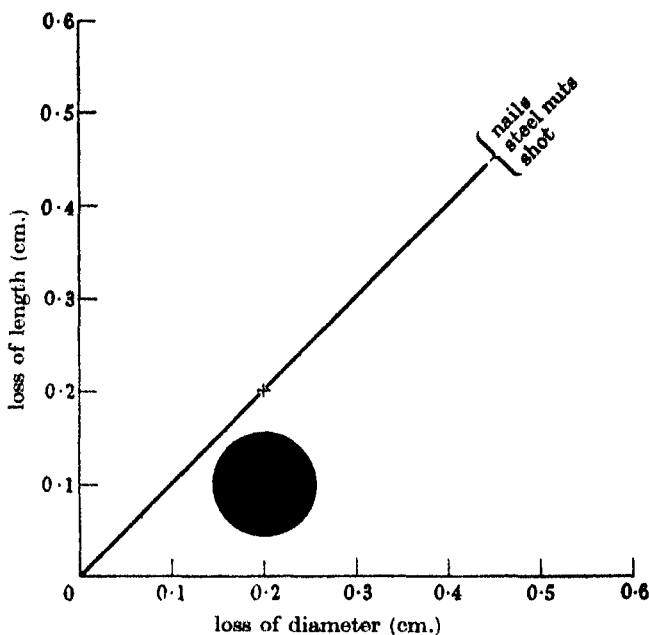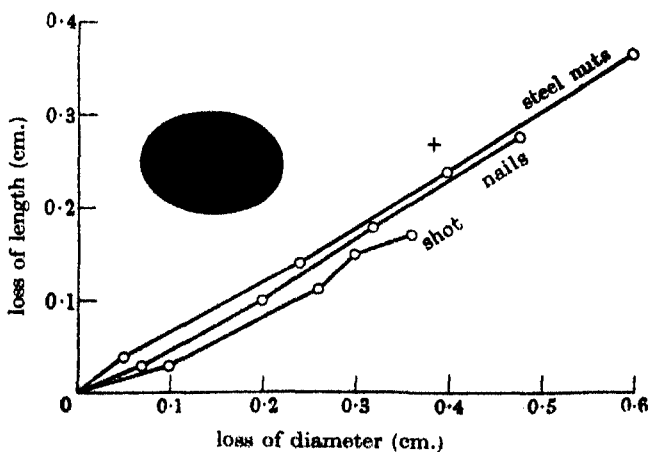

FIGURE 9. Spherical.

FIGURE 10. Oblate.

The box rotated in a lathe about a central axis parallel to the intermediate dimension, and the contents fell from one end to the other about once a second. The chalk pebble was taken out from time to time, and measured axially, and along three or (preferably) two equatorial diameters, by sliding calipers.

In some experiments chalk spheres of various sizes and chalk spheroids were abraded together for comparison, but no further reference is made to these tests in the present paper.

If a glass lid is placed on the tin box it becomes possible to observe the pebble during attrition, and to see if there is any marked tendency for it to assume one presentation or another towards the stream of impinging abrasive. So far as could be seen, there was no marked tendency of this kind, whether the pebble was prolate or oblate.

The experimental results are plotted on the diagrams, figures 7–12. It will be observed that the losses of length (polar) and diameter (equatorial) are plotted, instead of the actual lengths and diameters. These losses have to be determined by the difference of much larger quantities, which is disadvantageous, and the uncertainty (say, 0·02 cm.) is appreciable on the scale of the diagrams. The irregularities could be smoothed out by making adjustments of this order of magnitude.
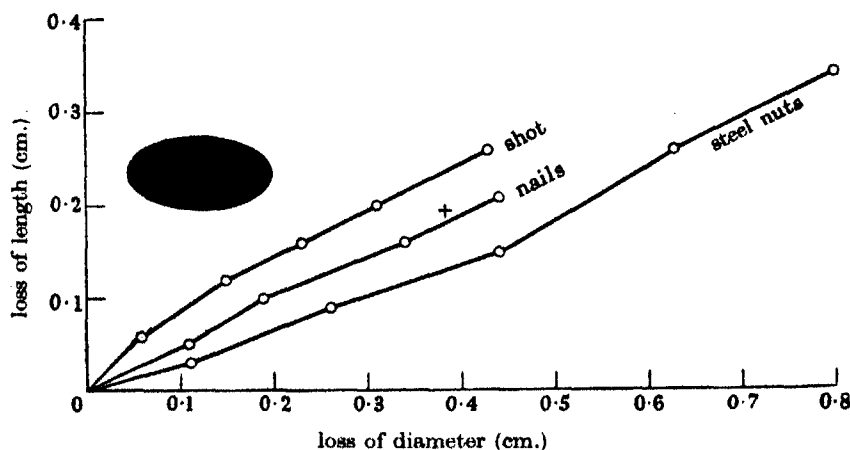
FIGURE 11. Oblate.

The diagrams are arranged in order, with the loss of diameter plotted as abscissa, and the loss of length as ordinate. Thus the axis of figure of the spheroid is to be thought of as vertical in each diagram as generally throughout this paper. The spheroids pass from prolate through spherical to oblate along the series of diagrams; the sections inset in each show these in comparative dimensions. On each diagram there is a point marked thus ×. The abscissa of this point is (on the scale of the diagram) 1/10 the original diameter of the spheroid, and its ordinate 1/10 given on the original length. Joining this point to the origin we should get a line whose

slope indicates the original ratio of length to diameter. (The line is not actually drawn because it would tend to overcrowd some of the diagrams.) I refer to it as the standard line.

If the ratio length/diameter were unaltered by abrasion, the graphs would be straight lines coincident with the standard line, the cross × lying on this line into which the graphs all coalesce. This would of course be actually realized for the sphere, the standard line being at 45°. The spherical case is indicated in figure 8, which merely expresses the fact that a sphere remains a sphere whatever kind of abrasive materials are used. This has often been incidentally verified in the present work, though verification is hardly needed. The only object of figure 9 is to show the relation with other cases at a glance.

It will be observed that in certain of these other cases the ratio length/diameter remains unaltered within the limits of experimental error. This applies, e.g. to prolate spheroids with the ratio length/diameter = about 1·4 when the abrasion is by steel nuts (figure 8).

It also applies to oblate spheroids, length/diameter about 0·5, abraded by nails (figure 11).

In all remaining cases shown on the diagrams the abrasion definitely disturbs the ratio of axes, and for prolate spheroids the ratio length/diameter is diminished. In other words the prolate spheroid either maintains its shape, or is shortened towards the spheroid form.

An oblate spheroid may either:

(1) Maintain its ratio of length/diameter as in figure 11 (nails),

(2) or (more commonly) the ratio length/diameter may increase, tending towards the sphere,

(3) or the ratio length/diameter may decrease, as in figure 11 (shot), length/diameter 0·5. This is a case unique in the diagrams of a tendency to change away from the sphere, the original oblate form becoming more oblate.

The changes of length and diameter are readily measured experimentally, and give a good deal of information: but clearly they cannot decide whether the figure remains spheroidal or not. For this purpose the projected shadow of the specimen must be examined.

The profile photographs (figures 13, 14a, 15, 16, 17, 18, 19, plate 2) illustrate the original shape (outer margin) and the final shape (inner margin). Figures 14b and c show the components of figure 14a separated. The composite photographs were made as follows. The specimen was tacked to a glass plate with sealing wax, and the photographic plate placed in contact with this glass plate. A distant arc lamp threw a sharp shadow of practically actual size. A similar photograph of the brass stencil used to shape the spheroid originally was made. Positives were printed by contact on glass from these two negatives, assembled film to film in the proper relative positions, and fastened together with sealing wax. The pictures here given were printed from these assemblages.

(It was at first attempted to adjust the chalk pebble in the centre of the stencil,

and to take the original picture at one exposure, but a satisfactory adjustment was not easily attainable in this way.)

It appears very clearly from these photographs that the figure does not in general remain spheroidal. A spheroid, whether prolate or oblate, is flattened at the poles away from the spheroidal form, as in the case of the natural pebbles in figures 2 and 3. This is evident by inspection of the photographs, and has been verified more objectively by the method illustrated in figure 1, though this is hardly necessary. Figures 13, 14 and 15 are prolate figures. The remainder are oblate.

The flattening of the prolate spheroid is at the place where the initial curvature and rate of abrasion is greatest. The flattening of the oblate spheroid is at the place where the initial curvature is least, and the linear rate of abrasion small. (It is to be noticed however that this rate of abrasion is not the minimum, which is at an intermediate latitude. Its precise location is difficult.)
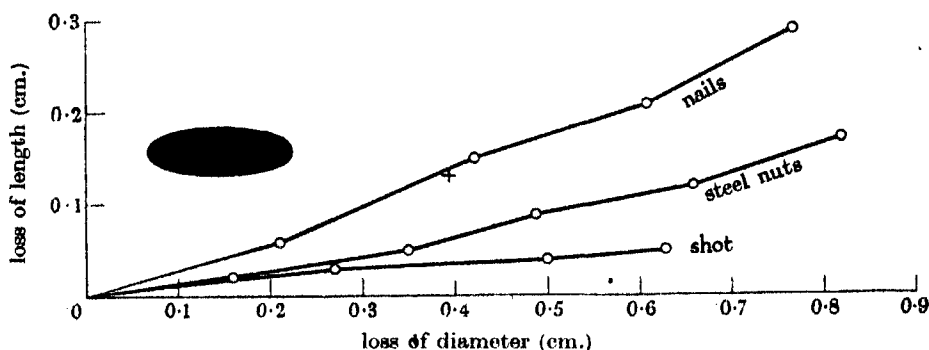
FIGURE 12. Oblate.

Consider the case where the ratio of the axes is preserved unaltered as in figure 13 (see also figures 8, 11, 12). In such a case, as we have seen, the abrasions at the pole and at the equator must be as the fourth roots of the initial specific curvatures. If the same held everywhere, the spheroidal form would be preserved.

As it is *not* preserved, we must conclude that abrasion is not a function of the specific curvature only. The same follows from the fact that starting from a spheroid of given shape and therefore given curvatures at the poles and the equator, we find in general a different ratio of polar to equatorial abrasion according to what abrasive is used—nails, steel nuts, or small shot. It is therefore useless to propose an amended law connecting the specific curvature with the linear abrasion at that point, for no such law can hold good.

This raises the question what the linear abrasion at a point on the surface *does* depend upon, if it is not determined by the specific curvature there. The considerations above given about the different abrasives show equally that it cannot be fixed by any other function of the principal curvatures, their sum for example. What then remains? The edges of a rectangular block are quickly rounded, which shows that a large local *principal* curvature (as distinguished from the specific

curvature) has an important influence, but it cannot be the only one. The general outline of the specimen must have an effect in guiding the motion of the abrading bodies large or small in its neighbourhood, and hence the shape at points other than the one under consideration must be important.

In most cases it would be difficult to attempt any further analysis; but there is a special case which lends itself to the attempt. This is the case of a flat surface with sharp edges. If abrasion depended solely on curvature, the flat surface ought not to be abraded at all; but common sense rejects such a conclusion, and this alone makes it evident that curvature is not everything.

Experiments on flat surfaces in fact lead to a very curious result. If we reduce an oblate ellipsoid sufficiently we get an approximately flat surface as in figure 19. After this amount of reduction, the specimen of chalk becomes somewhat rough and irregular though it shows a tendency to concavity. To pursue the matter further, it was found desirable to start with a flat surface, and to use steatite rather than chalk because it is of more uniform hardness and does not develop these local irregularities.* Starting with a flat disk of steatite with or without rounded edges, and abrading it with nails or steel nuts, a definite concavity developed on the initially flat faces, and shows conspicuously when tested with a straight-edge. Tests made with templates indicate that the radius of curvature of the concavity is of the order of 20 cm. This is seen in the photograph, figure 20, where a straight-edge placed over the surface shows the concavity by diffused light passing below, as with the concave natural pebble, figure 6. This concavity can be equally well obtained using as abrasive rounded flint pebbles instead of the steel nuts; it has therefore no necessary connexion with the angularity of the latter. On the other hand it is not obtained with small shot as the abrasive. We may therefore base the interpretation on assuming that the abrasive masses are comparable in size with the specimen.
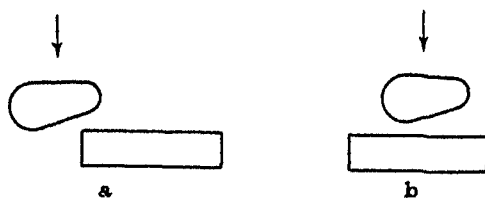
FIGURE 21 *a, b*. Illustrating formations of a concave surface.

Suppose now that an abrasive mass approaches the specimen, and let its path be such that the centre lies beyond the edge at the point of impact (figure 21*a*). The edge will be rounded, and the abrasive mass will be deflected away, and will not produce any effect on the flat surface near the edge. If, on the other hand, the centre of abrasive mass strikes definitely inside the edge (figure 21*b*) it will dig

* Probably steatite would have been a better material to use throughout, but I was for a time unable to procure it.

a hole in the flat surface at the point of impact. Thus, it is fairly easy to see that the statistical result will be a rounding of the edge, and the digging of a concavity in the flat surface.

It will be noticed that this explanation requires the impinging abrasive mass to be of considerable size. The explanation applies whether the impact is normal or moderately oblique.

It is believed that flint pebbles such as figure 6, after passing the flat stage have become concave in this way. It should be remarked that concave flint surfaces can originate by fracture, and the general shape and patination sometimes indicate that this is in fact their origin. There is no absolutely certain criterion for deciding, but the indications are that a flat biconcave specimen like figure 6 has been abraded concave.

In the experiments on abrasion which have been described, the abrasion has not been carried to great lengths, the object being rather to study the initial rate of abrasion of a spheroid. The tendency is usually towards a sphere, but it is so slow that the pebble will in most cases have disappeared before the sphere is attained. A few experiments have been made when the attrition was carried to great lengths. Thus, a rectangular block of chalk was cut measuring $2.5 \times 2.5 \times 1.6$ cm. and was ground down using steel nuts as the abrasive. The following results were obtained:

weight, g.	approximate shape	length/diameter
16·3	rectangular	0·694
4·0	oblate spheroid	0·71
0·09	oblate spheroid	0·74

Reducing the weight 183-fold has only achieved small progress towards the sphere.

Some tests were made on a prolate ellipsoid of chalk, abraded by steel nuts:

length cm.	diameter cm.	ratio
3·66	2·92	1·26
2·51	1·96	1·28

It will be seen that over this range there is little change of relative dimensions, and what tendency there is is away from the sphere. When the abrasion was carried further, the circular symmetry was lost, the form becoming ellipsoidal rather than spheroidal. If this is typical it accounts for the rarity of prolate pebbles of circular symmetry.

We do occasionally find spherical pebbles however, and they are often rather good approximations to a perfect sphere. It would seem therefore that when a certain approximation to a sphere has been attained, progress towards equality must be much more rapid than the above tests would suggest as likely. There is much more to be done in this direction, and materials harder and more homogeneous than chalk should be tried.

APPENDIX

Proof that if an ellipsoid is found to be reduced by abrasion without change of shape, the thickness removed at any point P must be as the fourth root of the specific curvature at P.

Consider the quadrics confocal with the ellipsoid, and passing through P. One of these will be a hyperboloid of one sheet, the other a hyperboloid of two sheets. Let the semi-axes co-linear with the semi-axis a of the ellipsoid be a' and a'' respectively. Let p be the perpendicular from the centre on to the tangent plane at P.

Then we have (Salmon 1914, p. 171)

$$p^2 = \frac{a^2b^2c^2}{(a^2-a'^2)(a^2-a''^2)}.$$

The principal curvatures at the point P are (Salmon 1914, p. 159)

$$p/(a^2-a'^2) \quad \text{and} \quad p/(a^2-a''^2).$$

Thus for the specific curvature we have

$$p^2/(a^2-a'^2)(a^2-a''^2) \quad \text{or} \quad a^2b^2c^2/(a^2-a'^2)^2(a^2-a''^2)^2.$$

The ratio of this to the specific curvature at the end of the semi-axis a is

$$\frac{a^2b^2c^2}{(a^2-a'^2)^2(a^2-a''^2)} \div \frac{a^2}{b^2c^2} \quad \text{or} \quad \frac{b^4c^4}{(a^2-a'^2)^2(a^2-a''^2)^2}.$$

If the material removed by abrasion at P has the thickness δp then, since similarity is assumed to be preserved, the ratio of this to the abrasion δa at the end of a is given by

$$\frac{\delta p}{\delta a} = \frac{p}{a} = \frac{1}{a} \times \frac{abc}{(a^2-a'^2)^{\frac{1}{2}}(a^2-a''^2)^{\frac{1}{2}}} = \frac{bc}{(a^2-a'^2)^{\frac{1}{2}}(a^2-a''^2)^{\frac{1}{2}}}.$$

This is the fourth root of the ratio of specific curvatures above found.

REFERENCES

Bullows, W. L. 1939 *Quarry Manager's J.* 21, 277–279.
Krumhein, W. C. 1941 *J. Geol.* 44, 482–520.
Salmon, G. 1914 *Geometry of three dimensions*, 6th ed. London and Dublin.
Showe, W. H. 1932 *Am. J. Sci.* 224, 111.
Thomson, W. & Tait, P. G. 1879 *Treatise on natural philosophy.*
Wentworth, C. K. 1919 *J. Geol.* 27, 507.
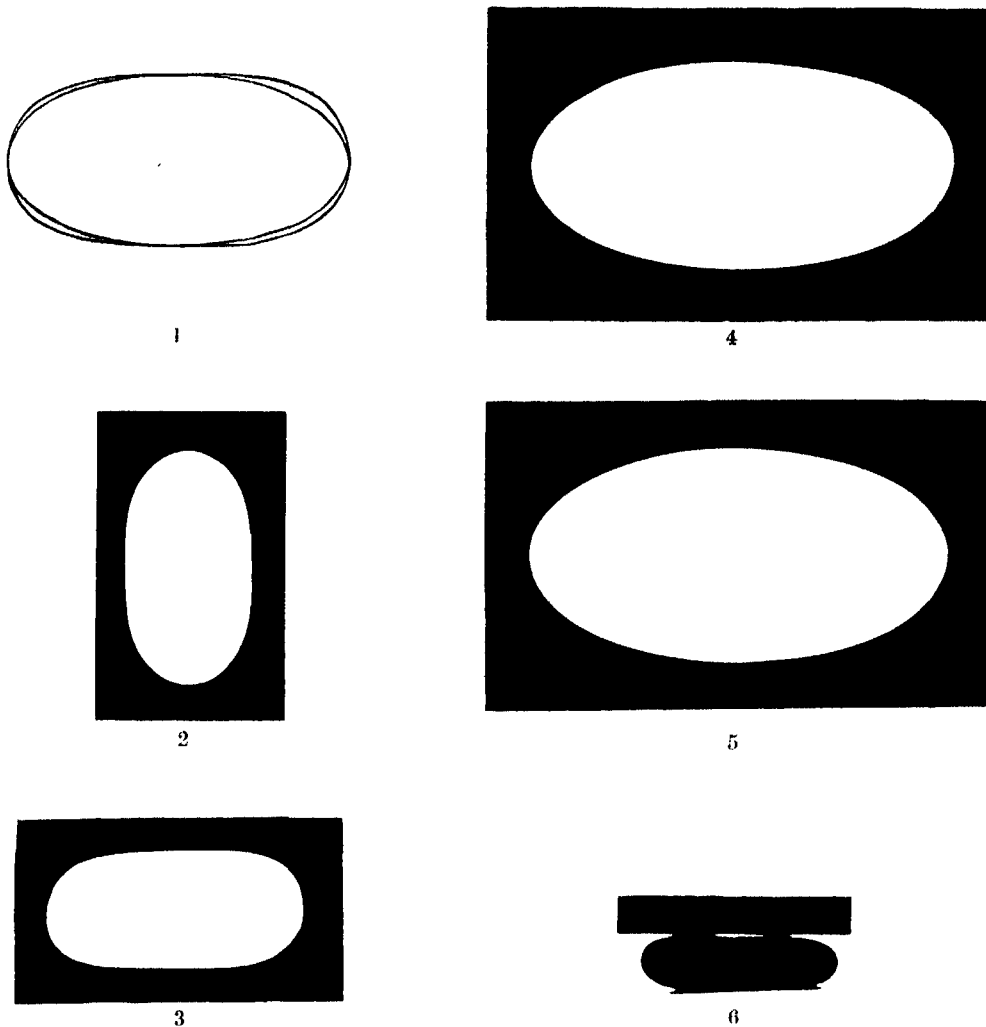
FIGURES 1–6. Natural pebbles. Actual size. Shown in section, axis of figure vertical.

FIGURE 1. The *ac* section of a quasi-ellipsoidal pebble. Ellipse inscribed.
FIGURE 2. Prolate quasi-spheroidal flint pebble. Axis vertical.
FIGURE 3. Oblate quasi-spheroidal flint pebble. Axis vertical.
FIGURE 4. Oblate spheroidal quartzite pebble. Axis vertical.
FIGURE 5. Ellipse of nearly the same dimension as figure 4, for comparison.
FIGURE 6. Biconcave pebble of flint. A straight-edge placed over it. Diffused light passes below the edge.
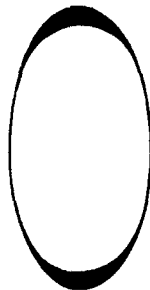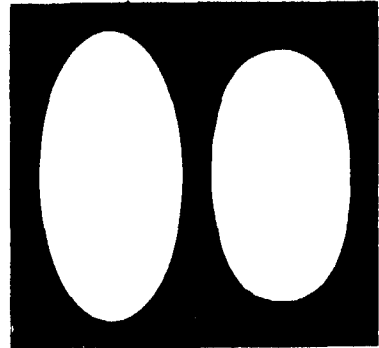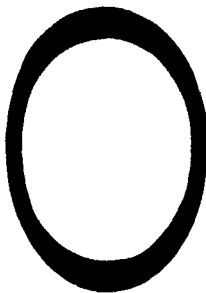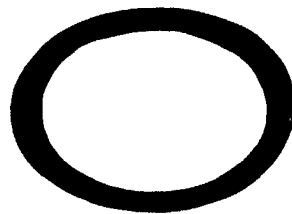
FIGURES 13–19. Experimental pebbles, spheroids of chalk.
Outside shows initial outline, inside shows final outline. Axis vertical. Actual size.

FIGURE 13. Prolate. Abraded with nails.
FIGURE 14*a*. Prolate. Abraded with small shot.
FIGURE 14 *b, c*. The same, with the initial and final figures separated.
FIGURE 15. Prolate. Abraded with nails.
FIGURE 16. Oblate. Abraded with nails.
FIGURE 17. Oblate. Abraded with nails.
FIGURE 18. Oblate. Abraded with steel hexagon nuts.
FIGURE 19. Oblate. Abraded with nails.

FIGURE 20. Steatite plane, abraded concave. Straight-edge applied as in figure 6.

Freezing-points of solutions of typical colloidal electrolytes; soaps, sulphonates, sulphates and bile salt

By Stewart A. Johnston and James W. McBain, F.R.S.

On the basis of experience with different families of colloidal electrolytes, each of which exhibits a whole range of typical behaviour different from each other, and in agreement with the experience of Hartley and his collaborators (1937), it is emphasized that conductivity alone is an untrustworthy guide in many cases as to whether or not a particular solution is that of a colloidal electrolyte. In all such cases it is essential to adduce other physical chemical evidence. Most conclusive is the direct comparison of thermodynamic with electrical properties.

Careful freezing-point studies have been made with a number of colloidal electrolytes. It is shown that the behaviour of different groups in solution falls into several different types, although all have in common the replacement of ions by colloid particles with increase in concentration. For example, in the family of bile salts, as in those of certain wetting agents, the conductivity almost approaches the behaviour of an ordinary electrolyte, whereas the lowering of freezing-point falls off strongly and rather abruptly.

It is pointed out that the term 'critical concentration' of micelles, as often used, is either an over-simplification or a misconception. Micelles are actually formed over a fairly wide range of concentration, amounting to at least 10-fold.

It is found that, in accordance with previous measurements of vapour pressure by McBain & O'Connor (1940) that loading potassium oleate solutions with solubilized iso-octane scarcely affects the lowering of freezing-point, showing that it is not solubilized as independent molecules, but that it is taken up in or upon existing colloidal particles.

Colloidal electrolytes were first recognized and defined by one of us in 1913 by comparing the osmotic behaviour of their solutions with their electrical conductivity. However, in recent years most investigators have omitted consideration of osmotic properties and have developed their conceptions from the electrical properties of modern synthetic solubilizing detergents. It is therefore important to supply measurements for a few typical cases. We have used for this purpose freezing-point equipment modelled after that of Scatchard (see McBain, Dye & Johnston 1939).

Experimental[*]

Potassium oleate

Potassium oleate is the most soluble higher soap. However, like all other soaps derived from the fatty acids, it is hydrolysed in dilute solution, unless a slight excess of potassium hydroxide is present. It has been shown that this excess remains as such (McBain & Martin, 1914; McBain, Taylor & Laing 1922).

Hence, potassium oleate was made from Kahlbaum's purest oleic acid, of iodine number 90 and molecular weight 288·2, with 4 equivalents % excess of potassium

[*] Measurements by S. A. J.

[119]

hydroxide prepared from potassium and water vapour free from carbon dioxide. This solution was stored in a bottle of Jena glass with a well-vaselined stopper to exclude carbon dioxide.

In the calculations the effect of the potassium hydroxide was calculated on the basis of the Debye-Hückel theoretical effect for that particular concentration. Since the alkali is not appreciably sorbed by soap (McBain & Martin 1914; McBain *et al.* 1922), we have made a good first approximation as to its effect. The theoretical depression due to the alkali was subtracted from the actual effect, and the remainder, which is due to the potassium oleate itself, used to calculate the value of the osmotic coefficient g. This is defined as $g = 1 - j = e/(1 \cdot 858 \times 2m)$, where e is the observed lowering of freezing-point in degrees, and m is the molality or weight normality, N_w.

We have found it possible to study the freezing-point down to solutions as dilute as $0 \cdot 00043$ molal KOl ($+ 0 \cdot 0000172m$ KOH), where the depression is only $0 \cdot 00157°$. The results are given in table 1 and in figures 1 and 2. The last three results were obtained quite independently by Mr A. P. Brady, using a different standardization and a solution made from another Kahlbaum's oleic acid of iodine number $90 \cdot 1$ and molecular weight $290 \cdot 4$.

TABLE 1. FREEZING-POINT LOWERINGS OF DILUTE AQUEOUS SOLUTIONS OF POTASSIUM OLEATE CONTAINING 4 EQUIVALENTS % EXCESS OF POTASSIUM HYDROXIDE

conc.	θ	$g_{corr.}$	conc.	θ	$g_{corr.}$
0·0004302	0·00157	0·981	0·01862	0·00963	0·1391
0·0007798	0·00262	0·904	0·02405	0·01102	0·1233
0·001198	0·00358	0·805	0·02901	0·01238	0·1149
0·002201	0·00485	0·593	0·04007	0·01569	0·1050
0·007253	0·00726	0·269	0·0998	0·0385	0·1039
0·009136	0·00773	0·228	0·1228	0·0473	0·1037
0·01165	0·00830	0·192	0·1625	0·0817	0·1022

In figure 1, the straight line from the upper left-hand corner represents the Debye-Hückel slope $g = 1 - 0 \cdot 32m$ for fully dissociated uni-univalent electrolytes in high dilution.

Only five values for freezing-point lowering of potassium oleate were previously available (McBain, Laing & Titley 1919). Two, shown in figure 2, were $0 \cdot 215°$ and $0 \cdot 348°$ for $0 \cdot 1$ and $0 \cdot 6$ N, corresponding to values of g of $0 \cdot 144$ and $0 \cdot 156$ respectively. These values by the Beckmann method appear to be substantially correct. Titley, by the Richards method, measured the three further dilutions, which were unfortunately progressively in error (the first, for $0 \cdot 206$ N_w, misprinted as $1 \cdot 196$ N_w, gave a g value of $0 \cdot 153$; and the others even higher values). A. P. Brady has found, for a $0 \cdot 366m$ potassium oleate with 4 equivalents excess of potassium hydroxide by the Beckmann method, three values of g, corrected for the excess potassium hydroxide, $0 \cdot 126$, $0 \cdot 135$, and $0 \cdot 134$, as shown in figure 2.
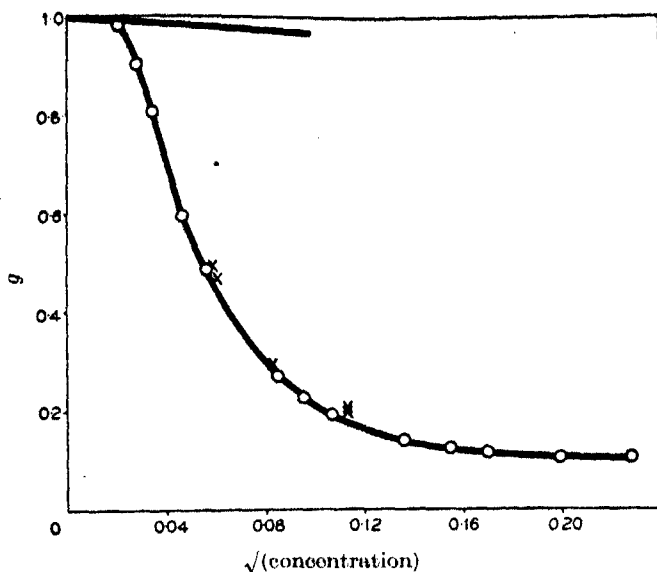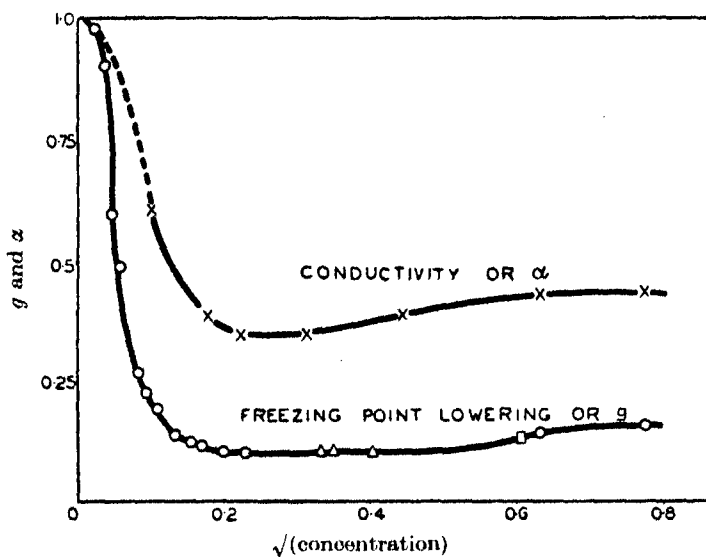
FIGURE 1

FIGURE 2

Sodium oleate

For sodium oleate, there have been only the two measurements (McBain *et al.* 1919) by the Beckmann method and two (Laing & McBain 1920) by the dew-point method at 18° and the graphs (without numerical data or tables) of Roepke & Mason (1940) by the thermoelectric vapour-pressure method at 25°. Mr O. E. Bolduan

finds that the sodium oleate made from metallic sodium and pure Lapworth oleic acid, supplied by British Drug Houses, with equivalent weight 285·0 and iodine value 85, has a solubility at 0° of not much over 0·01 N_w, where g would be approximately 1 anyway on account of the great dilution. It was therefore necessary to study supersaturated solutions, using 4 equivalents % excess of sodium hydroxide to prevent hydrolysis. Mr Bolduan failed to obtain a successful measurement with 0·4 N_w sodium oleate, and doubt is thrown upon the previous Beckmann value for this concentration (Laing & McBain 1920). The new procedure included the important precaution of inspecting the solution immediately after making the measurement to see if it was indeed clear and transparent and free from either curd or acid soap. His results for 0·2 N_w sodium oleate confirm the previous Beckmann value (McBain *et al.* 1919). Thus we have for pure sodium oleate at 0°, $g = 0·127$ for 0·2 N_w with some hydrolysis, as compared with the new results by O.E.B. of 0·104, 0·118, mean 0·111 for 0·2 N_w without hydrolysis. At 18° the dew point (Laing & McBain 1920) had given, as might be expected, the somewhat higher value of 0·17 for 0·4 and 0·6 N_w oleate. Thus sodium oleate is as osmotically active as potassium oleate; in both cases g would appear to pass through a minimum, although the evidence depends partly upon comparison between different methods of measurement. The values by Roepke & Mason, namely, 0·3 for 0·2 N and 0·25 for 0·4 N at 25°, are unaccountably high.

Aqueous potassium oleate with iso-octane solubilized therein

Detergents in dilute aqueous and non-aqueous solution solubilize otherwise insoluble substances. In spite of much speculation and conflicting opinions, no physical chemical evidence, apart from the X-ray measurements of Kiessig & Philippoff (1939), has been published regarding the actual mechanism of solubilization.

In the first place, a series of investigations (Krishnamurti 1929; Philippoff & Hess 1937; Hess, Kiessig & Philippoff 1938; Hess *et al.* 1939*a*; Hess, Philippoff & Kiessig 1939*b*; Stauff 1939*a, b*) have independently shown that clear transparent unhydrolysed soap solutions and solutions of soap-like detergents give good X-ray diagrams. These X-ray diagrams demonstrate that the colloidal particles are highly organized in a manner somewhat similar to crystals of soaps or fatty acids. The soap molecules are arranged in planes or layers or lamellae. The soap molecules are in pairs standing end to end, and these pairs are laid parallel to each other normal to the layers. Hess and his collaborators regard these lamellar micelles as being truly crystalline, whereas Stauff (1939*a, b*) regards them as being liquid crystalline, in that the molecules may be free to rotate on their individual axes.

Kiessig & Philippoff's contribution (1939) was to show that the spacing, 49 A, corresponding to the double length of the sodium oleate molecule, is increased to 85 A when benzene was solubilized therein. This indicated that a layer of benzene,

36 A thick, is sandwiched between the single layers of close-packed soap molecules separating the hydrocarbon ends. In addition, they found layers of water 42 A thick sandwiched between the polar side of the soap layers making a total spacing of 127 A. Stauff's 'Grossmizellen' are thus identical with the neutral or lamellar rather poorly conducting micelle proposed by one of us many years ago. In more dilute solutions in which colloid is still shown to be present by conductivity and by freezing-point these X-ray diagrams cease to appear, showing that smaller less organized micelles, such as an ionic micelle, are the first to be formed. Nearly all authors now recognize that at least two kinds of colloidal particles or micelles are formed by any one colloidal electrolyte.

In Lester Smith's (1932) studies of the solubilization of aniline or of a dye, he concluded that 'the solvent powers of soap solutions can only be accounted for by postulating adsorption of the organic solute on the colloidal soap particles'. Hartley and Lawrence, on the contrary, later suggested that dyes are dissolved in a liquid interior of globular micelles. We consider that, in addition to the Kiessig & Philippoff apparently proven case, we may also have to postulate yet further mechanisms, such as colloidal particles containing both soap and solubilized material, to account for some of the large colloidal particles which become evident in the ultramicroscope when the soap is not too concentrated nor insufficiently loaded. Whether, then, the exterior is mostly soap or whether Lester Smith's conception is a better description remains an open question.

Ward (1939, 1940), studying sodium dodecyl sulphate in water and aqueous alcohol and noting that alcohol is insoluble in paraffin, concludes that it is sorbed on the exterior of the micelles to the extent of between one and four molecules of alcohol per equivalent of sulphate in the micelle.

In any case, the results we now present establish the fact that the solubilized molecules do not exist as such; that is, they do not dissolve as free molecules, but are wholly taken up in colloid form, preferably in or on existing soap micelles.

This was previously shown by the fact that the volatile hydrocarbon disappears into the soap solution in large quantity without producing a high vapour pressure; and we now show similarly, and as a necessary corollary, that it dissolves in the soap solution without appreciably increasing the lowering of freezing-point. Conversely, the addition of the *iso*-octane cannot greatly alter the colloidal electrolyte.

Table 2 gives values for solutions of the same potassium oleate as in table 1, but where now *iso*-octane was added in nearly saturation amount, about 0·7 g. to 1 soap. Since the vapour pressure of the hydrocarbon is well below saturation, no emulsion can be present, but only solubilized material (McBain & O'Connor 1940). The results are compared in figure 1 (points marked by crosses) with those for the solution without *iso*-octane, and it is seen that the effect of the *iso*-octane is negligible. Had the added *iso*-octane remained in the form of ordinary separate independent molecules, each fully active osmotically, the value of g reckoned on the

soap alone would have been raised very greatly, namely, by no less than 0·98, whereas practically no effect is seen. Thus in the strongest soap solution the *iso*-octane alone, if in true solution, would have produced five times the observed lowering.

TABLE 2. FREEZING-POINT LOWERING OF POTASSIUM OLEATE SOLUTIONS NEARLY SATURATED WITH SOLUBILIZED *iso*-OCTANE

m, KOl	θ	g	m, KOl	θ	g
0·003335	0·006149	0·4961	0·01275	0·009452	0·1996
0·003494	0·006126	0·4719	0·01278	0·009600	0·2021
0·006796	0·007438	0·2946			

Sodium decyl sulphonate

Results have already been published for the non-hydrolysing detergent, dodecyl sulphonic acid (see McBain *et al.* 1939). Its sodium salt proved too insoluble to be investigated at 0°. Sodium decyl sulphonate is somewhat more soluble, and results for the available range of concentrations are given in table 3 and figure 3. The purified salt was kindly supplied by Professor H. V. Tartar. It is of interest to note that in the figure, the very dilute solutions almost coincide with the Debye-Hückel values, the most dilute giving a value within 0·1% thereof. At 0·017m the value of g begins to fall rapidly.

TABLE 3. FREEZING-POINT LOWERING OF SODIUM DECYL SULPHONATE

conc.	θ	g	conc.	θ	g
0·001931	0·007086	0·987	0·01083	0·03853	0·958
0·003157	0·01151	0·982	0·01554	0·05408	0·936
0·004911	0·01779	0·975	0·01724	0·05965	0·931
0·007126	0·02574	0·972	0·01835	0·06127	0·899
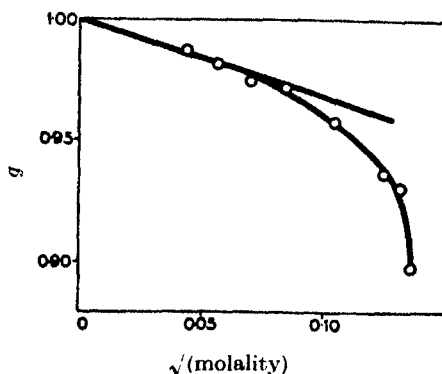
FIGURE 3

Sodium decyl sulphate

The sulphate likewise does not hydrolyse, although it is more soluble and less stable if acidified. We are indebted to the research laboratories of the E. I. du Pont de Nemours and Co. for purified specimens of this and the dodecyl sulphate. The measurements are given in table 4 and figure 4. The somewhat high results must be attributed to remaining impurity, as shown also by analysis for sodium sulphate. The osmotic coefficient falls rapidly above $0.026m$.

TABLE 4. FREEZING-POINT LOWERING OF SODIUM DECYL SULPHATE

conc.	θ	g	conc.	θ	g
0·001551	0·0057	1·00	0·007614	0·0280	0·99
0·002037	0·0075	0·99	0·01118	0·0417	1·00
0·002743	0·0101	0·99	0·01371	0·0506	0·99
0·003854	0·0134	0·97	0·02461	0·0895	0·98
0·005142	0·0189	0·99	0·03705	0·1232	0·89
0·005837	0·0213	0·98	0·05016	0·0363	0·73
0·006878	0·0256	1·00			
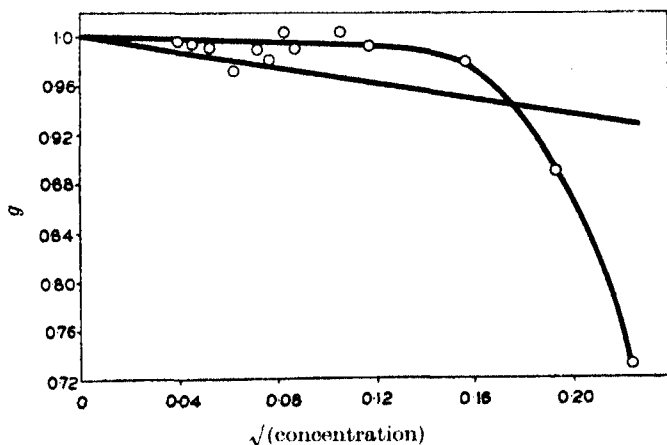
FIGURE 4

Sodium dodecyl sulphate

With the dodecyl sulphate the osmotic coefficient begins to leave the Debye-Hückel slope well before its rapid descent at $0.005m$. At 25°, where it is very soluble, the conductivity at this concentration is still following the Debye-Onsager slope (McBain & Searles 1936). Insolubility prevents following it further. Our results are displayed in table 5 and figure 5.

TABLE 5. FREEZING-POINT LOWERING OF SODIUM DODECYL SULPHATE

| conc.	θ	g	conc.	θ	g
0·0007519	0·00276	0·987	0·002537	0·009115	0·967
0·001022	0·00374	0·984	0·002920	0·01032	0·962
0·001231	0·00450	0·984	0·004039	0·01420	0·946
0·001764	0·00640	0·976	0·004776	0·01672	0·942
0·002093	0·00753	0·968	0·005485	0·01798	0·882
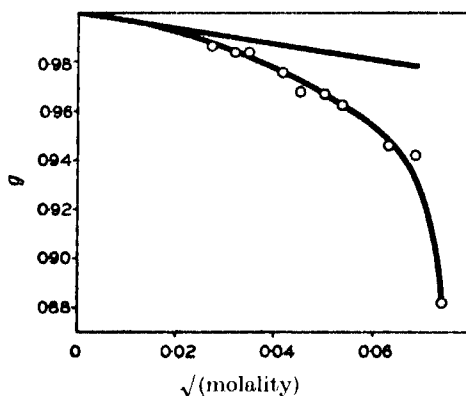
FIGURE 5

The bile salt, sodium deoxycholate, as a colloidal electrolyte

The bile salts, such as sodium deoxycholate, $C_{23}H_{37}(OH)_2.COONA$, are sterol derivatives and colloidal electrolytes as has been shown also by the contemporaneous work of Roepke & Mason (1940), who have given some data for the cholate, glycocholate and glycodeoxycholate. Our sodium deoxycholate was specially supplied by Riedel-de Haen. It was dried at 105° for 1 hr., losing 10 % of its weight, presumably as water vapour. Our measurements of freezing-point extend from 0·003 up to 0·31m. The osmotic coefficient drops rapidly between 0·007 and 0·07m, after which it continues to fall slowly without passing through a minimum. The results are shown in table 6 and figure 6.

TABLE 6. FREEZING-POINT LOWERING OF SODIUM DEOXYCHOLATE

conc.	θ	g	conc.	θ	g
0·002984	0·0111	1·00	0·02794	0·0650	0·63
0·003397	0·0124	0·98	0·02959	0·0672	0·61
0·003817	0·0141	0·99	0·04054	0·0824	0·55
0·004349	0·0162	1·00	0·05772	0·105	0·49
0·005638	0·0208	0·99	0·07358	0·124	0·45
0·007412	0·0267	0·97	0·1024	0·159	0·42
0·01029	0·0352	0·92	0·1430	0·208	0·39
0·01313	0·0411	0·84	0·2128	0·286	0·36
0·01555	0·0456	0·79	0·3235	0·409	0·34
0·02183	0·0556	0·68			

For comparison, the conductivity was carefully measured (*cf.* McBain *et al.* 1939) at 0° C with a cell whose constant was 12·517. The values are given in table 7 and plotted in the form of conductivity ratio in figure 7. The equivalent conductivity at infinite dilution would be 37 mhos by extrapolation, or 36 mhos if in more
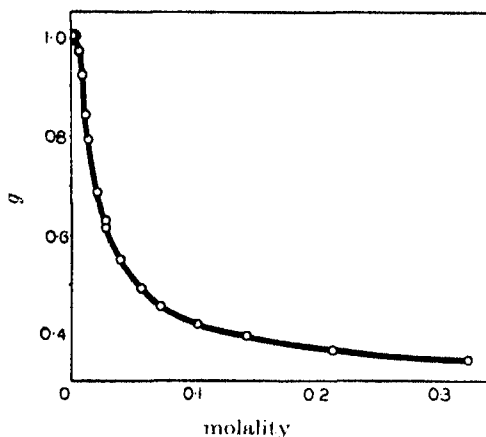
Figure 6

dilute solutions the curve changes to the Debye-Hückel slope. With increasing concentration the conductivity, quite unlike that of a higher soap or the higher sulphates or sulphonates, merely falls steadily over the whole range of concentration, as shown in table 7. Similar behaviour was shown by Roepke & Mason for their three bile salts. This is in striking contrast to the behaviour of the osmotic coefficient which is like that of the soaps and sulphates and sulphonates.

TABLE 7. CONDUCTIVITY OF SODIUM DEOXYCHOLATE AT 0° C

conc.	equiv. conductance	conc.	equiv. conductance
0·02129	34·96	0·07065	30·77
0·03284	33·83	0·09926	29·25
0·03934	33·18	0·1335	27·72
0·05789	31·87	0·2949	22·48

DISCUSSION

The primary fact about all colloidal electrolytes is that, with sufficient increase in concentration, ions are gradually replaced by colloid particles. This substitution always takes place over a wide range of concentration. Thus, the critical concentration for completion of the formation of micelles is ten or often many more times greater than the 'critical concentration for the initiation of micelle formation'.

The effect of substituting colloidal particles or micelles for ions is usually greatly to reduce the equivalent conductivity and the lowering of freezing point, although in different degree. This is the most striking effect to be seen in all the figures of this paper. No other adequate explanation for it has yet been suggested.

Different colloidal electrolytes differ quite markedly from each other in the relations of conductivity and osmotic behaviour in dilute solution. They differ also in comparison with the Debye-Hückel formulations for fully ionized electrolytes. It is, therefore, misleading to take any one, for example, such as the alkyl sulphates or sulphonates studied by Lottermoser, Hartley and many others, as typical. Indeed, as has been known for over 20 years, in each family, even within homologous series, such as the original soaps or the sulphonic acids, the behaviour changes gradually over the whole range from lower to higher members. The bile salts, and several groups of wetting agents, differ from both of these families, not so much as regards osmotic behaviour, but in respect to conductivity. Their conductivity undergoes no great diminution at any definite concentration, but falls off only slowly and gradually over the whole range.

Even the dyes differ from each other. Methylene blue is so far unique in having a short range of conductivity rising to a value much greater than that at infinite dilution before it begins to fall off. However, the higher paraffin chain sulphonic acids exhibit a conductivity (McBain et al. 1939) that in extreme dilution does not fall off with concentration according to the Debye-Onsager slope for strong electrolytes, but remains constant or horizontal above it before beginning the steep descent. A similar behaviour was reported by Tartar for related magnesium sulphonates.

Hartley and collaborators (1937) quote their previous results with methylene blue, and their very different results with meta benzopurpurin whose conductivity simulates that of an ordinary strong electrolyte, and state explicitly: 'This means that it may be impossible to obtain from conductivity measurements alone direct evidence of the formation of ionic micelles'. This *caveat* has been overlooked in several other laboratories in recent studies.

Bile-salt type

Hence, we may discuss just two of these types. First, let us take the latest to be examined, namely, the bile salts. Figures 7 and 8, respectively, give our data for sodium deoxycholate at 0° and those of Roepke & Mason (1940) for sodium glyco-deoxycholate at 25°, drawn on the same scale. It is seen that they are very similar in type and they resemble the cholate and glycocholate (Roepke & Mason 1940). For convenience in displaying the curves, *g* and equivalent conductivity are both expressed as unity at infinite dilution.

With the four bile salts, the ions rapidly disappear without much affecting the conductivity. Therefore, the colloidal particles must be small, highly charged micelles, not like those in ordinary concentrations of soap. It is impossible to determine without such further evidence as migration, on X-ray evidence, whether

the micelles are all of one kind, produced from a mixture of ions, or whether there are present some ionic micelles composed essentially of anions. There can be no important concentration of ordinary ion-pairs or of molecules, for in such case the conductivity would be more affected than the osmotic behaviour.
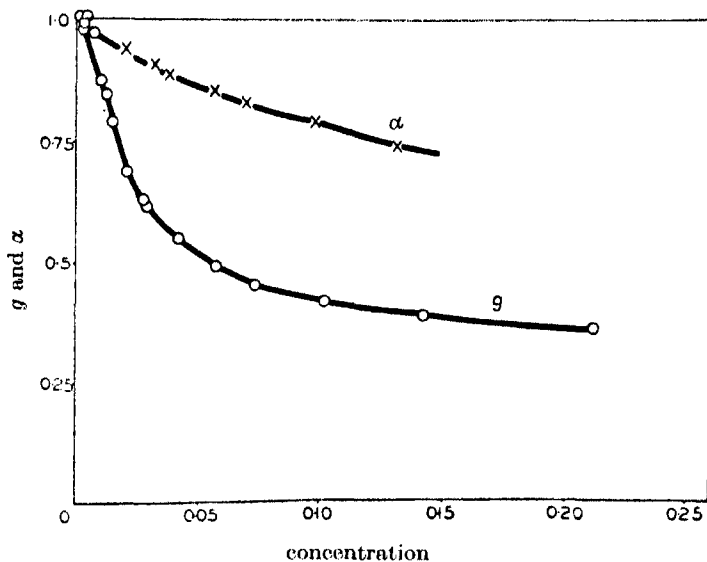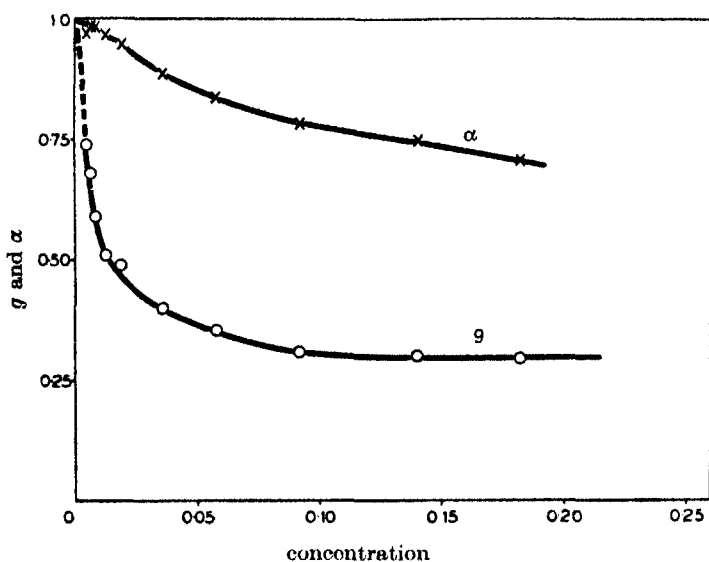

FIGURE 7

FIGURE 8

Higher soap and lower homologous types

For the soaps, we have the new data for the oleates. These were collected in figure 2 for potassium oleate. In this, values of freezing-point lowering expressed as g and values of conductivity (McBain *et al.* 1919) (at 18°) expressed in terms of that at infinite dilution are plotted against the square root of molality, or weight normality. Both g and conductivity have, therefore, the value of unity at infinite dilution.

Figure 2 shows that the so-called critical concentration for the initiation of micelles, if there be such a thing, comes near 0·0004 N, but that the lowering both in conductivity and in osmotic coefficient is not completed until about 0·06 N; that is, a range of concentration of 150-fold, during which the proportion of colloid progressively increases. Next, it is seen that both conductivity and osmotic activity are very greatly lowered. Conductivity is much less lowered, only to one-third, whereas osmotic activity passes through a definite, although shallow, minimum, rising again with further increase in concentration by 12 and 50 %, respectively. This is also unmistakable for the lower potassium soaps, data for which are appearing elsewhere. For the osmotic effect this must be due to hydration, whereas for conductivity McBain has ascribed it to progressive formation of highly conducting ionic micelles. Hartley and others formerly departed from the principle of Le Chatelier and of mass action which were relied upon in more dilute solution, and assumed that dissociation of the associated ions now increases instead of decreasing with concentration. Ward (1939, 1940) queries this, and concurs in Hartley's (1939) more recent attempt to explain it in terms of overlapping ionic atmospheres and changes in dielectric constant of water extending to comparatively great distances from the micelles.

McBain (1940) points out that the absolute concentration of ions must necessarily increase throughout, and that formation of the small ionic micelles will likewise increase throughout, beginning at extreme dilution. Again, from the same principle, the formation of the large lamellar micelles of different formula, earlier given the unfortunate designation of 'neutral' colloid, must increase relatively very fast as soon as its formation begins until it is fairly complete, even if of modified composition and size and, therefore, of shifting equilibrium constant.

Stauff's lamellar colloidal particle is identical in properties and structure with McBain's lamellar or 'neutral' micelle. Hartley's 'ionic micelle' of a few years ago was similar in conductivity with McBain's neutral micelle, but different from the latter in not being lamellar and not changing in size and composition with concentration. Ward (1939, 1940), using conductivity alone and ignoring the X-ray evidence for lamellar micelles from several other laboratories, retains a spherical, poorly conducting or relatively neutral micelle like Hartley's, with the properties of McBain's lamellar micelle. The radius of the sphere is fixed by the chain length which makes it independent not merely of concentration but of nature of solvent. McBain always regarded this lamellar micelle as the most important colloid present,

as it was necessary to account for the great *lowering* in both conductivity and osmotic coefficient. The highly conducting ionic micelle was brought in to obviate quantitative discrepancies between a whole group of physical chemical properties.

Van Rysselberghe (1939), taking into account both our conductivity and freezing-point data, has shown that the average composition of the micelle varies with concentration over an extremely wide range beginning with a simple ion-pair. Mrs M. E. L. McBain (1933) has shown that the diffusion of the oleates is quantitatively in accord with the interpretation in terms of lamellar along with ionic micelles. This is likewise the case for her striking results for free lauryl sulphonic acid (1939).

The enhanced conductivity in very dilute solutions referred to above, where conductivity is greater than the Onsager limit for complete dissociation, for several types of colloidal electrolytes must be due to simple, highly conducting micelles, such as McBain's ionic micelles, and not to ion-pairs or even groups of two anions with one cation. Comparison of conductivity and freezing-point or dew-point lowering for the *lower* soaps (compare the following section) indicates (McBain 1939) that neutral ion-pairs are the first and most prominent products of association over a relatively wide range of concentration. There is no alternative to this conclusion if there is only one product of association and if the conductivity ratio $\alpha = 2g - 1 =$ van't Hoff's $i - 1$, more especially when both α and g and i are corrected for Debye-Hückel effects.

It has been repeatedly pointed out that the free electrical charges on all micelles proposed by McBain, Hartley, Stauff, etc. are so far apart that they are practically independent and each has an independent ionic atmosphere (compare Kirkwood 1934). Hence the ionic strength is that of a uni-univalent electrolyte. Consonant with this is the repeatedly established fact (for example, McBain & Searles 1936) that both conductivity and freezing-point lowering or vapour-pressure lowering are altered to the same extent by additions of indifferent electrolytes as if they were measured in the presence of uni-univalent salts. For this reason, conductivity and thermodynamic properties of colloidal electrolytes may be interpreted as if they referred to uni-univalent electrolytes. To avoid misunderstanding, it should be emphasized that the indifference of conductivity and osmotic coefficient of colloidal electrolytes to added salts is only observed after the colloid is fairly completely formed already. Indeed, like Powney & Addison (1937) and others, we have found that in extreme dilution added salt has a great influence in forming colloid. Here added salts increase solubilizing action, whereas in less dilute solution they tend to diminish it.

Turning again to figure 2 for potassium oleate, it follows that since the conductivity curve lies so much higher than the osmotic coefficient curve, the average compositions of the colloid must correspond to a high proportion of paired off positive and negative ions and yet the soap is fairly highly conducting. Were this to be referred to one kind of colloid with that conductivity, the migration number so predicted would be far greater than that observed (Laing 1924). Therefore, we are

forced to place most of the neutral soap in the poorly conducting lamellar micelles and retain only a moderate proportion of highly conducting small ionic micelles consisting largely or wholly of associated ions of one kind.

Only gradual and small effects are observed with lower homologues

It is interesting to recall, in connexion with all this work, the data for the lower soaps for comparison with the other figures in this paper. This is done in figure 9. Comparison of these with the impressive series of measurements by Lottermoser & Püschel (1933; sodium dodecyl sulphate alone has also been carefully studied by Howell & Robinson 1936, and by Ward 1939, 1940), for the salts of the higher alkyl sulphates, shows that it is for the highest homologues in the straight-chain sulphonates and sulphates that there is the most abrupt drop in both conductivity and osmotic coefficient in rather dilute solution, whereas for the lower homologues in all cases the falling off is very gradual, becoming important only at really high concentrations such as N/2 or higher. A so-called critical concentration is an extreme case confined to the highest homologues of a very few types of colloidal electrolytes. Bolam & Hope (1941) point out that phenanthrene sulphonic acids and their derivatives resemble the lower homologues of the paraffin-chain derivatives.
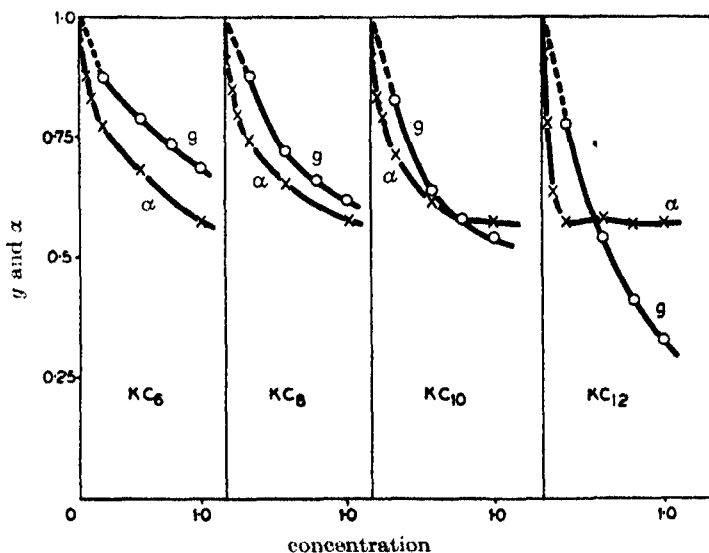
FIGURE 9

Ward (1939, 1940) finds a gradual transition from colloidal electrolyte to normal electrolyte on addition of alcohol in progressively large quantities to aqueous sodium dodecyl sulphate. He states: 'From 40 % of alcohol onwards there is no critical concentration' and concludes that 'the fraction of the ions which aggregate to micelles at the critical concentration is gradually decreased and falls to zero at about 40 % of alcohol'.

REFERENCES

Bolam & Hope 1941 *J. Chem. Soc.* p. 843.
Hartley 1939 *Kolloidzschr.* 88, 22.
Hartley, Collie & Laurie 1937 *Trans. Faraday Soc.* 32, 796.
Hess, Kiessig & Philippoff 1938 *Naturwissenschaften*, 26, 184.
Hess, Kiessig & Philippoff 1939a *Naturwissenschaften*, 26, 184.
Hess, Philippoff & Kiessig 1939b *Kolloidzschr.* 88, 40.
Howell & Robinson 1936 *Proc. Roy. Soc.* A, 155, 386.
Kiessig & Philippoff 1939 *Naturwissenschaften*, 27, 593. For a copy of their diagrams see
    Advances in colloid science, p. 124, Interscience Publishers Inc., N.Y. (1942).
Kirkwood 1934 *J. Chem. Phys.* 2, 351.
Krishnamurti 1929 *Indian J. Phys.* 3, 307.
Laing 1924 *J. Phys. Chem.* 28, 673.
Laing & McBain, J. W. 1920 *J. Chem. Soc.* 117, 1513.
Lottermoser & Püschel 1933 *Kolloidzschr.* 63, 175.
McBain, J. W. 1939 *J. Phys. Chem.* 43, 677.
McBain, J. W. 1940 *Nature, Lond.*, 145, 702.
McBain, M. E. L. 1933 *J. Amer. Chem. Soc.* 55, 545.
McBain, M. E. L. 1939 *Proc. Roy. Soc.* A, 170, 415.
McBain, M. E. L., Dye & Johnston 1939 *J. Amer. Chem. Soc.* 61, 3210.
McBain, Laing & Titley 1919 *J. Chem. Soc.* 115, 1289.
McBain & Martin 1914 *J. Chem. Soc.* 105, 969.
McBain & O'Connor 1940 *J. Amer. Chem. Soc.* 62, 2855.
McBain & Searles 1936 *J. Phys. Chem.* 40, 493.
McBain, Taylor & Laing 1922 *Trans. Chem. Soc.* 121, 621.
Moilliet, Collie, Robinson & Hartley 1935 *Trans. Faraday Soc.* 31, 120.
Philippoff & Hess 1937 *Ber. dtsch. chem. Ges.* 70, 1808.
Powney & Addison 1937 *Trans. Faraday Soc.* 33, 1253.
Roepke & Mason 1940 *J. Biol. Chem.* 133, 103.
van Rysselberghe 1939 *J. Phys. Chem.* 43, 1054.
Smith 1932 *J. Phys. Chem.* 36, 1675.
Stauff 1939a *Naturwissenschaften*, 27, 213.
Stauff 1939b *Kolloidzschr.* 89, 224.
Ward 1939 *J. Chem. Soc.* p. 522.
Ward 1940 *Proc. Roy. Soc.* A, 176, 412.

An electrical detector of condensation in high-velocity steam

By A. M. BINNIE, M.A. AND J. R. GREEN, B.E., D.PHIL.

[Plate 3]

A new method has been developed for finding where condensation in a convergent-divergent steam nozzle commences. When a short length of fine wire, mounted on a rod, was traversed axially through the nozzle, it was found that its resistance altered sharply at a certain point. The position of this point agreed closely with that of a small sudden pressure rise in the nozzle, which an earlier investigation had shown to indicate the beginning of condensation of the supersaturated steam. The mean temperature of the wire upstream from this point was found (as predicted by Griffith's theory) to be nearly the same as that of the steam at entrance to the nozzle; it was greatly in excess of the temperature of the high-velocity steam passing the wire.

The wire was heated electrically, and the Nusselt number of heat transfer to the steam was measured under the prevailing supersonic conditions. The Wilson line, showing the position of condensation on the Mollier diagram, was determined, and is compared with the line previously obtained for another nozzle of different shape.

1. INTRODUCTION

A recent investigation (Binnie & Woods 1938) on the flow of steam through a convergent-divergent nozzle showed that the point of condensation was marked by a sharp rise in pressure. This method of detecting condensation can rarely be employed, and we have therefore sought for another which could be used in pipes and in turbine passages. It is well known that the various types of wind deflectors fitted to the bridges of ships and to the cockpits of aircraft become relatively ineffective when the air contains drops of moisture because these heavier particles are not easily diverted from their course. A similar principle underlies the design of centrifugal separators, and it has even been used in instruments for measuring wind-blown sand in the Sahara. It thus seemed probable that the action of flowing steam on the surface of a wire held in it would depend markedly on the presence or absence of moisture in the steam.

Accordingly, experiments were carried out on a modified form of the nozzle previously used. Being fitted with numerous pressure tappings, it served as a convenient device for producing steam which at a known and well-defined point became wet. A fine wire supported by a rod was arranged to move along the nozzle axis, and its resistance (and hence its mean temperature) was determined. In addition to the main object of the work, two subsidiary investigations were made: (*a*) the loss of heat from the wire at various temperatures was measured; (*b*) the pressure distribution in the nozzle was examined as in the previous tests, and it was then possible to show how far friction losses and the position of the Wilson line depend on the form of the nozzle.

2. Description of the apparatus

The apparatus, shown in figure 1, was a modification of that used in the earlier work. The rectangular nozzle previously employed was too shallow to permit the introduction of a wire. New cheeks were therefore inserted to form a deeper nozzle which had a somewhat greater throat area. The divergent walls were straight and inclined at 2·0° to the axis, the throat was formed by circular arcs, and the constant depth was 2·22 cm. Figure 2 gives a plan of this nozzle and indicates also the pressure tappings in the base; the dimensions are shown in table 1.
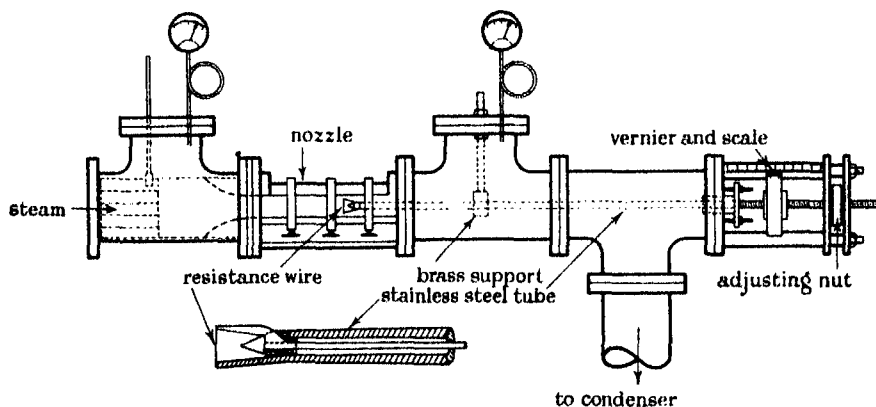
Figure 1. Arrangement of apparatus.
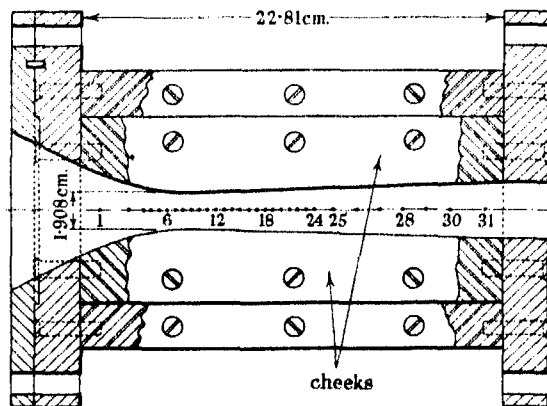
Figure 2. Plan of rectangular nozzle.

In the earlier work the steam pressures were read on independent mercury U-gauges, each with its own air leak and a leak indicator consisting of a water U-gauge. This arrangement was replaced by a multiple gauge (figure 3, plate 3), in which the twenty-five pairs of mercury and water columns rose from common reservoirs, supplied from tanks attached to the back of the gauge. The water reservoir was

A. M. Binnie and J. R. Green

TABLE 1. NOZZLE DIMENSIONS

tapping number	distance from tapping 1 cm.	mean measured width cm.	tapping number	distance from tapping 1 cm.	mean measured width cm.
1	·---	4·021	17	8·32	2·133
2	1·53	2·931	18	8·74	2·161
3	2·40	2·476	19	9·16	2·191
4	2·82	2·299	20	9·58	2·220
5	3·24	2·159	21	10·00	2·251
6	3·67	2·054	22	10·43	2·281
7	4·10	1·978	23	10·85	2·311
8	4·52	1·930	24	11·27	2·341
9	4·94	1·915	25	12·54	2·432
10	5·36	1·925	26	13·81	2·518
11	5·78	1·953	27	15·09	2·607
12	6·21	1·987	28	16·36	2·698
13	6·63	2·016	29	17·63	2·789
14	7·05	2·046	30	18·90	2·877
15	7·47	2·074	31	20·80	3·013
16	7·89	2·104	throat	5·04	1·908

divided into two parts; one served the leaks to which compressed air had to be admitted because the pressure to be measured approached or exceeded that of the atmosphere, while the other was connected to the leaks in the divergent part of the nozzle where the pressure was less than atmospheric. The pairs of columns at the ends of the row were left open so that they showed the levels in the mercury and water tanks, and thus gave the zeros for the other columns; the two mercury columns were also used as a level to set the scale on the gauge truly horizontal. The indications of the gauge were recorded by a 9×12 cm. plate camera with an exposure of $2\frac{1}{2}$ sec., the readings from the plates being corrected for parallax. This method of obtaining the pressures proved most successful. Much time was saved during the tests and a permanent record obtained. A total of 107 plates was usefully exposed.

As before, the apparatus was supplied with steam from a boiler through 50 ft. of $1\frac{1}{2}$ in. pipe, and the pressure at inlet to the nozzle was approximately atmospheric. In the earlier experiments the desuperheater and two throttling valves were fitted close to the apparatus, and in the course of the discussion on the paper several speakers suggested that the age of the steam at entry might be a factor of importance. An additional desuperheater was therefore fitted 10 ft. from the boiler, the water being introduced into the pipe by a centrifugal pump, and this gave the desuperheated steam at entry to the nozzle an age of about $\frac{1}{2}$ sec. The boiler stop valve was also used for throttling purposes, but it was found that whatever arrangement of throttling and desuperheating was tried the pressure distribution in the nozzle was unaffected.

The wire was carried on a tube, $\frac{1}{2}$ in. outside diameter and $\frac{3}{16}$ in. inside diameter, which (as shown in figure 1) protruded into the nozzle from the downstream end. The front end of the tube was tapered, so that immediately behind the wire its

outside diameter was $\frac{5}{16}$ in. A $\frac{1}{16}$ in. rod ran down the centre of the tube, and to its upstream end was brazed a conical cap, from which projected one of the prongs supporting the wire. The rod was insulated from the tube by beads spaced along its length, and its downstream end was fitted with a nut and terminal so that the rod could be tightened against bushes which at both ends insulated it from the tube. The second prong supporting the wire was brazed to the tube. To avoid possible thermo-electric and electro-chemical effects, all these fittings (apart from the bushes and the wire) were made of stainless steel. In addition to the rod, two fine copper galvano-meter leads were inserted into the tube. They were connected to the upstream ends of the rod and the tube, and in effect may be considered as attached to the ends of the wire. The tube was traversed axially by means of a nut, and its position was shown by a scale and vernier. The travel of the wire was restricted to the divergent part of the nozzle, where the stream velocity was supersonic; hence the flow upstream from the wire was unaffected by the presence of the wire and its support.
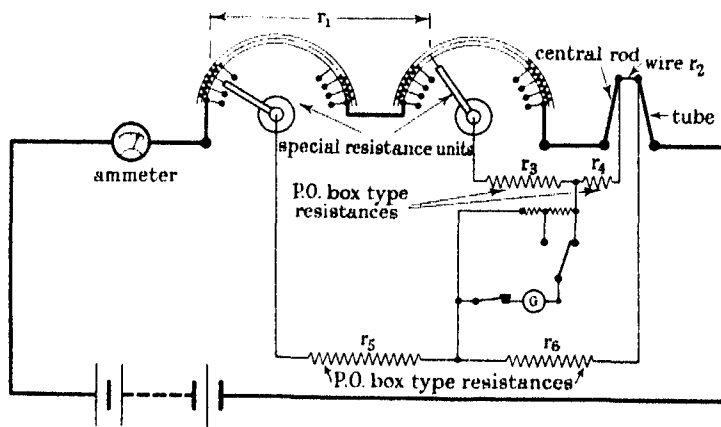
FIGURE 4. Arrangement of electrical apparatus.

In measuring the resistance of the wire it had to be borne in mind that the central rod (and to a lesser extent the tube) offered an appreciable resistance, which owing to temperature effects would vary in an incalculable manner. To avoid this difficulty the Kelvin double bridge (figure 4) was employed, in which r_3, r_4, r_5 and r_6 were fixed resistances with typical values 400, 20, 4000 and 200 Ω, and the resistance r_2 of the wire was about $0 \cdot 4 \Omega$. To facilitate the measurements a special variable resistance was built, consisting of a single length of 16 I.W.G. Eureka wire with a resistance of $11 \cdot 32 \Omega$. Two sets of tappings were provided, the intervals between the studs on the left dial being twenty times those on the right. In this way r_1 (and hence r_2) could be quickly determined without altering the current through the wire. The mirror galvanometer was connected through a potentiometer so that its sensitivity could be reduced when large currents were used. With this apparatus it was found possible to measure r_2 with an accuracy of about $0 \cdot 1 \%$.

A 42 I.W.G. nichrome wire was first tried as the resistance element, but it proved too fragile for the severe conditions in the nozzle. It was therefore replaced by a finer tungsten wire of 0·0514 mm. diameter, the ends of which were brazed to the steel prongs. Not only was its life satisfactory, but its temperature coefficient was 17 times that of nichrome, and it was therefore used throughout the investigation. The length of each of the tungsten wires used was 0·30 in. correct to 0·01 in., which was the limit of the accuracy of measurement because of the presence of a small blob of brazing material at the junctions with the prongs. The length/diameter ratio of the wires was about 150, so that end-effects may be ignored and the wire considered as a cylinder operating under uniform conditions throughout its length. It should be added that in all the tests the Reynolds number of the flow round the wire was far below the critical value for cylinders which is of order 10^5.

3. THE WIRE AS A DETECTOR OF CONDENSATION

In these tests the procedure adopted was as follows. The rod was first screwed as far downstream as possible, and the wire was then outside the nozzle. When conditions had become steady, the inlet thermometer was read and a plate exposed. The wire was then traversed up the nozzle, resistance measurements being made at intervals. This sequence of operations usually occupied about 8 min., during which period the inlet temperature rarely altered by more than 1° C.

The current through the wire was usually maintained at 80 mA. With larger currents the resistance and therefore the mean temperature of the wire began to
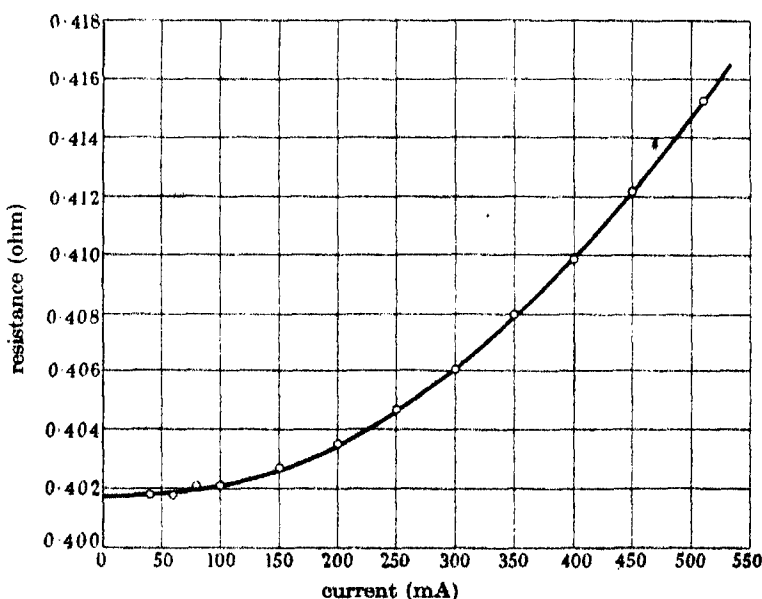
FIGURE 5. Current-resistance curve for the wire in supersaturated steam.
Test no. 118a; Reynolds no. = 456.
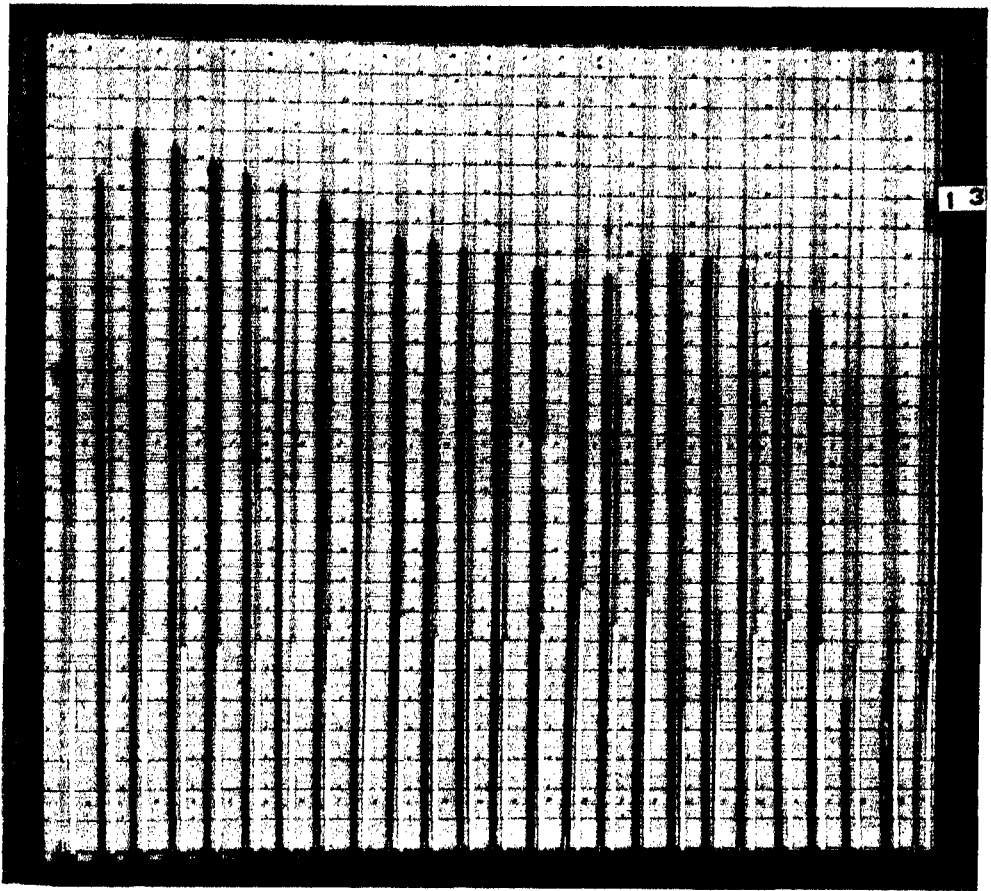
FIGURE 3. Multiple gauge.

rise, as is shown in figure 5. It was thought undesirable to complicate the issue by using the wire at a temperature appreciably exceeding that with zero current. The curve was obtained during a test by keeping the wire in one position in the supersaturated region, and measuring the resistance as the current was increased.

The results of six typical tests are displayed in figure 6, where the wire resistance and the ratio of the observed pressures P to the initial pressures P_0 are plotted on
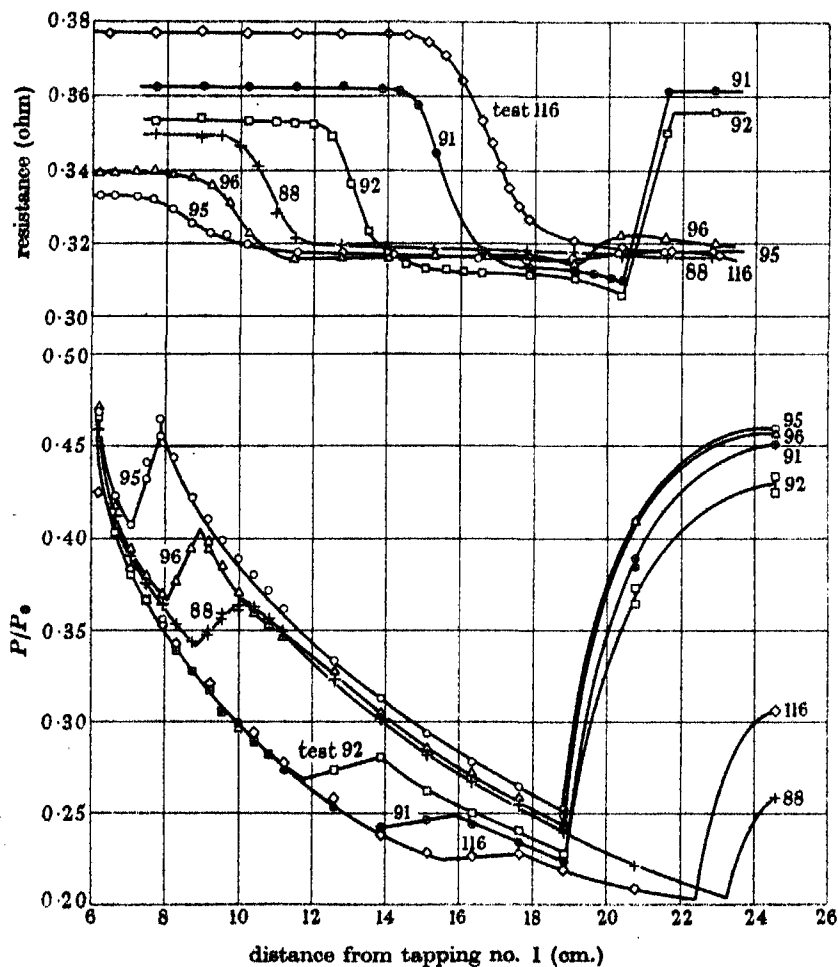
FIGURE 6. Resistance and pressure in the nozzle.

test number	P_0, lb./sq. in. abs.	T_0, °C
◯ 95	9·5	101
△ 96	9·6	108
+ 88	15·0	126
▢ 92	9·5	127
⊕ 91	9·6	136
◇ 116	14·0	153

a common base. It will be seen that the region of condensation corresponds to a well-marked fall in the wire resistance. In these experiments a recompression wave was present near the downstream end of the nozzle, and apart from tests 88 and 116 this occurred within the range of the wire measurements. In test 92 the resistance
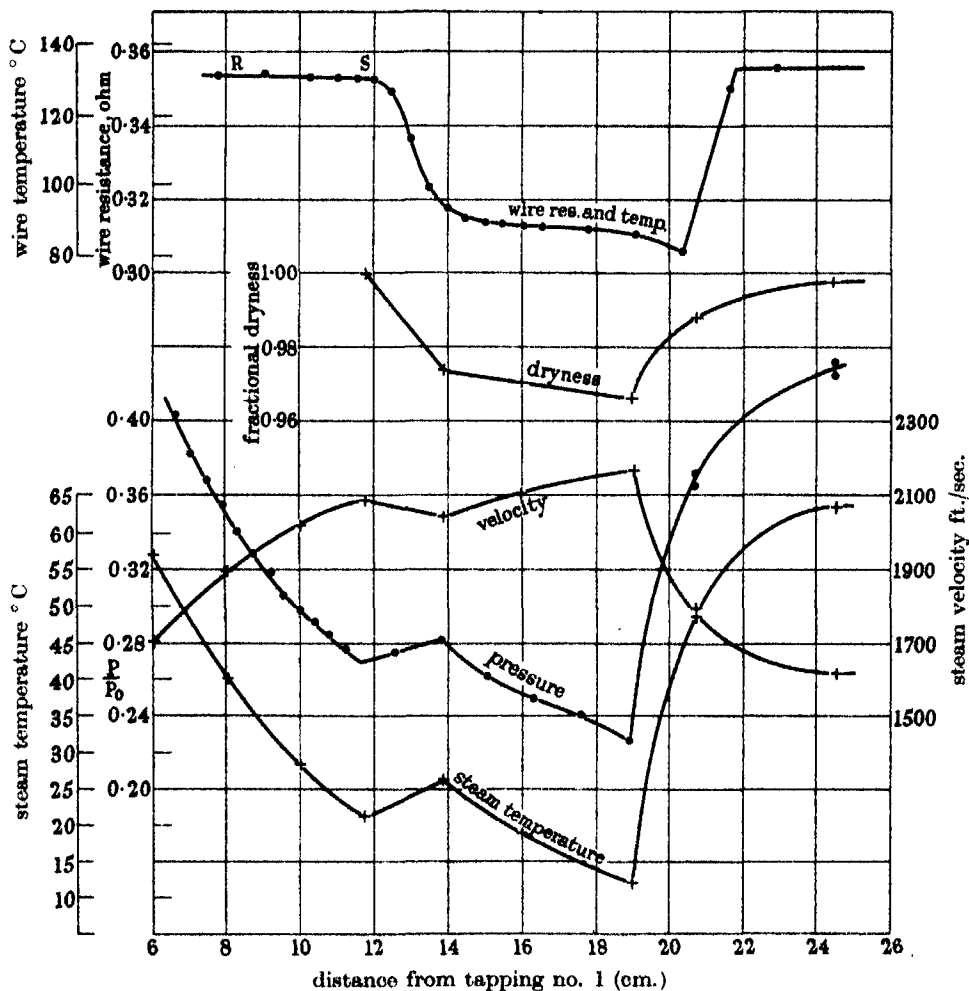
FIGURE 7. Analysis of test no. 92. $P_0 = 9.5$ lb./sq. in. abs.; $T_0 = 127°$ C.
● observed values; + calculated values.

curve in the region of the wave rose towards its initial value, and it is shown in the next paragraph that at the crest the steam was almost dry. Again, in test 91 with a greater initial superheat, the steam at the foot of the wave was drier and therefore at the crest was probably superheated; thus the corresponding increase in the resistance was even more marked. On the other hand, in tests 95 and 96 the steam was very wet at the foot of the wave, the drying process of recompression was in-

sufficient to remove all the moisture, and the rise in the resistance curves was slight. Therefore good evidence exists for believing that the resistance drop was due to the presence of moisture.

This conclusion was confirmed by a detailed analysis of test 92, the results of which are shown in figure 7. The temperature, velocity and dryness corresponding to a number of points on the observed pressure curve were approximately calculated, and reasonable curves drawn in. As far as the beginning of condensation the friction corrections were determined by the method explained in § 6, and in close agreement with the earlier work the condensation rise was taken as isentropic. A similar friction correction was applied in the remaining expansion. Although the recompression process involved a gain in entropy, for simplicity's sake it also was assumed isentropic; the resulting dryness was therefore underestimated. From the evidence of figure 7 it seems difficult to believe that the drop in resistance depended mainly on any other factor than the dryness of the steam. It will be noticed that (as in all the tests) the portion RS of the resistance curve was almost horizontal although the steam temperature was falling. This point is dealt with in the next section, where the question of the mean temperature taken up by an unheated conducting body is discussed.

FIGURE 8. Current-resistance curves (test 133).
+ in supersaturated steam, curve X. ○ in wet steam, curve Y.

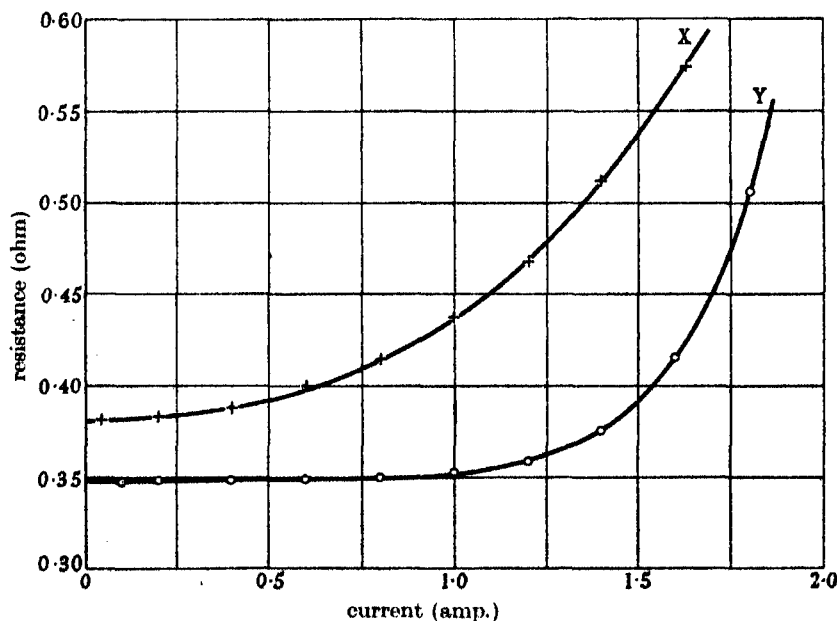Several tests, of which that shown in figure 8 is typical, were made to determine whether the current-resistance curves for the wire were different in the supersaturated and the wet regions. Comparatively large currents were used. Curve X

shows the result obtained with the wire in supersaturated steam: for Y the wire was held in wet steam downstream from the condensation pressure rise, the flow through the nozzle being the same for both. The difference in form of the two curves is very marked. In X the increase of energy supply to the wire led to a steady rise in temperature, while in Y the temperature rise was greatly delayed. It appears that in Y about $1\cdot0$ A was required to dissipate the effects of moisture, whether in the form of impacts of droplets or of a film on the wire, and not until this current was exceeded was it possible for the wire temperature to rise. The evidence of curve X, supported by figure 7, suggests that in supersaturated steam these effects were in comparison negligible.

One test was performed with steam initially so highly superheated that no condensation took place in the nozzle. When the wire was traversed up the nozzle, the resistance was found to be almost constant throughout, although it was estimated that over the last cm. of its travel the wire was in superheated steam. No change was noticed as the wire passed from the supersaturated to the superheated region.

4. THE TEMPERATURE TAKEN UP BY THE UNHEATED WIRE

The conditions which obtain round the surface of an infinite circular cylinder placed transversely in a stream of fluid vary greatly from point to point. It is therefore not to be expected that the mean temperature taken up by an unheated conducting cylinder can be closely predicted on theoretical grounds alone. Griffith, however, suggested in an unpublished paper that when compressibility effects are absent the temperature T_1 of the cylinder is related to the temperature T of the undisturbed stream by the formula

$$T_1 = T + v^2\{1 - (1-\sigma)\,m^2\}/2gJc_p, \qquad\qquad (4\cdot1)$$

where g is the acceleration due to gravity, J the mechanical equivalent of heat, v the velocity in the undisturbed stream, and c_p the specific heat (at constant pressure) of the fluid; m depends on whether the boundary layer is laminar or turbulent. The dimensionless Prandtl number $\sigma = \eta c_p/k$, where η and k are the viscosity and thermal conductivity of the fluid; it is here restricted to values of order unity. Equation (4·1) was experimentally verified by Hilton (unpublished), who used a high-speed wind tunnel and showed that the value of m was about $1\cdot57$. This held good until the Mach number $M = v/a$ (where a is the local velocity of sound) reached $0\cdot59$, where a discontinuity was observed in the otherwise linear relation between v^2 and the temperature difference. At this point the temperature of the cylinder suddenly rose, an effect which was attributed to the formation of a shock wave, which caused an increased amount of kinetic energy to be converted into heat. This explanation was confirmed by later experiments (Hilton 1938). Meissner (1938) also measured the temperature of high-speed gas flows and mentioned the work of Müller (1920) on the temperature distribution in the divergent part of a Laval nozzle, but this latter reference cannot at present be followed up.

Now the wire in the nozzle could be regarded not merely as a detector but also as a resistance thermometer, and it was calibrated to an accuracy of about $1°$ C by immersion in a well-stirred oil bath, the current being $40\,mA$. The relation between its temperature T_1 in degrees centigrade and its resistance r in ohms was found to be linear and given by $r = 0.2320 + 0.925 \times 10^{-3}\,T_1$ over the range $15-163°$ C, and this
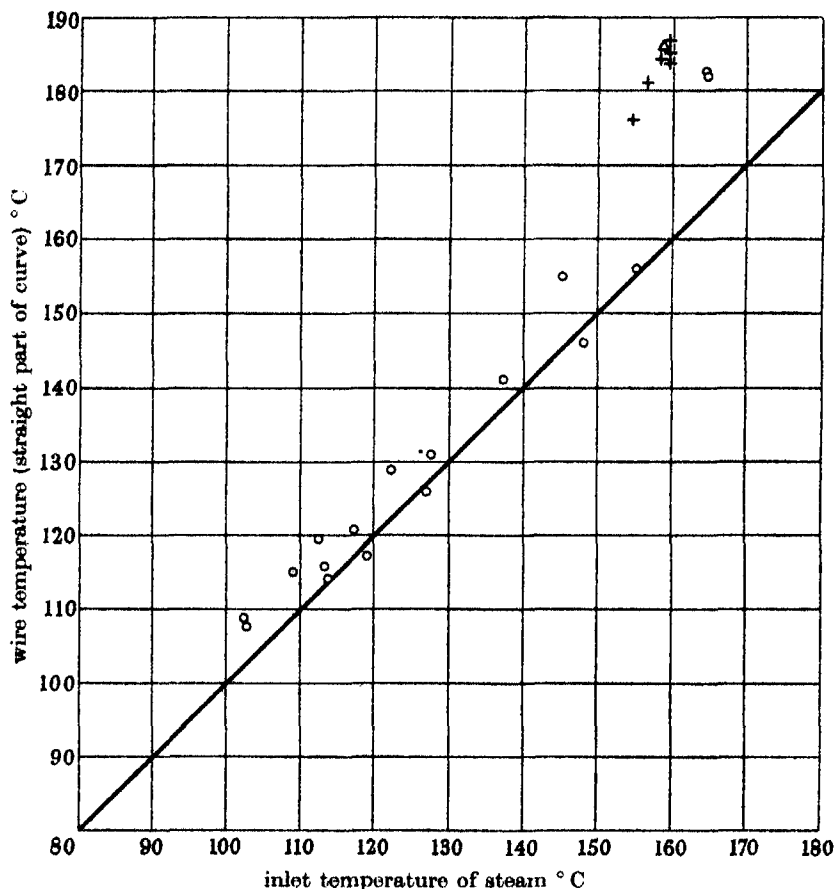
FIGURE 9. Comparison of wire temperature with inlet steam temperature.
○ supersaturated, supersonic; + superheated, subsonic;
△ superheated and supersaturated, supersonic.

result enabled the wire temperature graduations to be added to figure 7. In the same way it was possible to contrast the inlet steam temperature with the almost constant wire temperature observed in the nozzle upstream from the point of condensation. For twenty-five tests the comparison is shown in figure 9, which indicates that apart from a few scattered points the wire temperature exceeded the inlet steam temperature. The additional point marked 'superheated and supersaturated, supersonic' was obtained from the test mentioned at the end of § 3. Those marked 'super-

heated, subsonic' were derived by making the pressure difference between the nozzle inlet and outlet so small that no supersonic velocities were attained; they refer to single observations at definite positions in the nozzle.

To analyse the observations it was necessary to calculate the values of σ over the working range. The values of c_p were obtained from the formula given by Callendar (1920), k was taken from the observations of Milverton (1935) and η from those of Hawkins, Solberg & Potter (1935). But much extrapolation was necessary, and the results (shown in figure 10) cannot be regarded as highly accurate. However, it is apparent that in conformity with Griffith's assumption the value of σ was nearly unity, but unfortunately the form of (4·1) is such that the temperature difference $(T_1 - T)$ is very sensitive to small errors in the value of σ. The effects of variations in σ can be easily but crudely estimated from (4·1) if the simplifying assumptions are made that the flow was frictionless and c_p was constant during the expansion. Then T may be obtained from the inlet temperature T_0 by means of the equation

$$T_0 - T = v^2/2gJc_p, \tag{4·2}$$

which, when substituted in (4·1), yields

$$T_1 - T_0 = (T_0 - T)(\sigma - 1)m^2. \tag{4·3}$$

Thus if $\sigma > 1$ the wire temperature will exceed the inlet temperature. For the super-heated, subsonic tests in figure 9 the pressures at the wire were higher than in the supersaturated, supersonic tests, thus σ and the values of $(T_1 - T_0)$ obtained from (4·3) were greater, as is borne out by the observations.

Because of the doubtful accuracy of figure 10 detailed analysis was confined to test 116, in which (as shown in figure 6) the condensation pressure rise occurred far down, giving a long length of nozzle at nearly constant wire temperature. The conditions obtaining at four points in this region were calculated, and the results are given below in table 2. Friction corrections were inserted, and (4·3) was not used. The velocities were high not only actually but also relatively, as is indicated in line 4 which shows the values of the Mach number M. A comparison between lines 5 and 7 makes it clear that the temperature differences between the wire and the undisturbed stream were consequently very considerable. The wire temperatures suggested by Griffith's theory with $m = 1·57$ are shown in line 6. These temperatures lay just below the initial steam temperature 155° C, while the observed wire temperatures were just above it. It seems probable that these discrepancies were due to shock waves, for which no allowance could be made in the theory. It will be noticed that, in contrast with the preceding paragraph, the more refined method used here led Griffith's theory to predict wire temperatures below the inlet temperature, but with the prevailing large values of $(T_1 - T_0)$ and the uncertainty over σ it is perhaps surprising that such close agreement between theory and experiment was obtained.

These dynamical effects seem to have been overlooked in the earlier discussions (e.g. Martin 1918) on the temperature taken up by a thermometer bulb immersed in flowing supersaturated steam. No supercooling had been observed, as it was

found that the thermometer temperature was about the same as that corresponding to the saturation pressure. This was attributed solely to the deposition of a film of moisture on the bulb. Again, it has been suggested that one advantage of a Curtis wheel as the initial stage of a turbine is that although the steam may be supplied to the nozzles in a highly superheated state, its temperature has fallen greatly before it comes into contact with the blades and with the main body of the turbine case. Hence the temperature attained by these parts is much below that of the inlet steam. This conclusion is not supported by figure 9, and it should be added that the mean temperature taken up by an immersed body was shown by Meissner (1938) not to be dependent on its precise size.
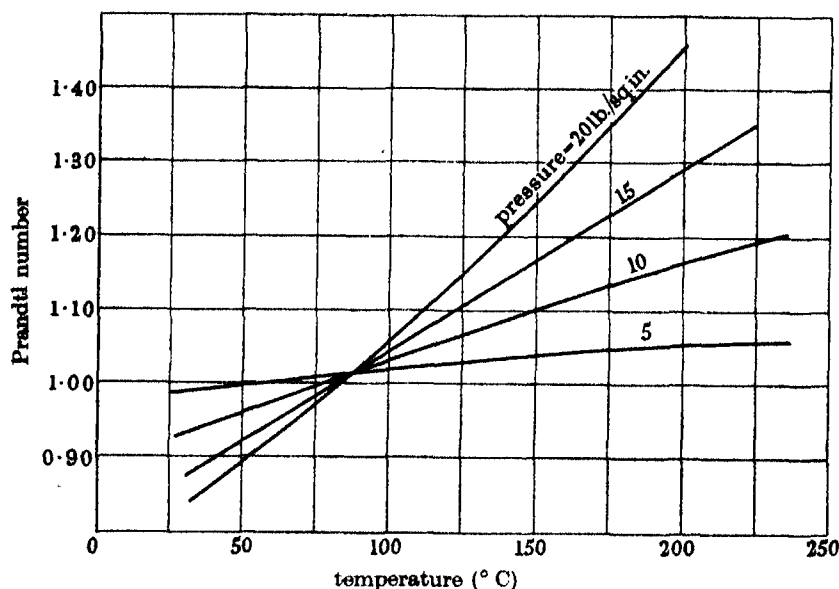
FIGURE 10. Prandtl numbers for steam.

TABLE 2. ANALYSIS OF TEST 116

Initial steam pressure 14·04 lb./sq. in., temperature T_0 155°C.
Wire current 0·080 A. $\sigma = 1·005$.

		A	B	C	D
1.	Point	A	B	C	D
2.	Distance from tapping 1 (cm.)	6·5	9·0	11·6	14·1
3.	Velocity v (ft./sec.)	1840	2030	2130	2230
4.	Mach number M	1·23	1·37	1·46	1·54
5.	Steam temperature (T°C)	73	55	44	34
6.	Wire temperature from (4·1) (T_1°C)	153	151	151	151
7.	Observed wire temperatures (°C)	157	158	157	157

---

here is the wire diameter d. Previous investigators of this subject, who all used fluids at low velocity, took the temperature of the undisturbed stream as the datum T. But in the present experiments, in which the temperature difference between the unheated wire and the undisturbed stream was so large, it seemed more correct to use the temperature of the unheated wire as datum. The value of k employed in (5·2) was that of the undisturbed stream.

In this way five curves, like that for supersaturated steam in figure 8, were analysed in an attempt to derive $f(R, M)$ for high-velocity conditions. They were obtained by measuring the current-resistance relation for the wire, while it was fixed in the supersaturated region with all other conditions constant; the steam pressures were read photographically in the usual manner. To estimate Nu, the energy input Q was plotted against the wire temperature. The results are shown in figure 11, from which it will be seen that the observations lay on straight lines, and from the slope of these Nu was calculated with the aid of (5·2). Finally, in table 3 the resulting values of Nu are given together with those of R, M and σ. The available information is too scanty to allow the form of the unknown function to be determined. Moreover, Nu could not be estimated with precision, and the values shown in column 2 are not correct in the third figure. Thus the table is best summarized by stating that over the small ranges of M and R which could be investigated the value of Nu was about 15. This may be compared with Hilpert's curve, which shows that $Nu = 12$ for $R = 500$. Again, the difference is due probably to shock waves, which may be expected to improve the heat conduction from the wire to the passing steam.

TABLE 3. VALUES OF NUSSELT NUMBERS

test	Nu	M	R	σ
117	15·4	1·65	434	0·985
118a	15·6	1·56	453	0·990
118b	15·1	1·48	499	0·995
118c	15·9	1·36	546	1·005
133b	14·7	1·20	550	0·994

Figure 12 has been drawn up to emphasize how the behaviour of the wire depended on the fluid in which it was immersed. The straight line A is taken from figure 11, while B is typical of several which (as mentioned in the second paragraph of §4) were obtained with superheated steam at subsonic velocities. A plate was exposed before the wire was moved from the position where the current-resistance curve was read, but, since the velocities were no longer supersonic, it was doubtful whether the pressure at the wire was the same as that shown on the photograph. Consequently, these results were not subjected to the analysis explained in the preceding paragraphs, but it can be asserted that the steam condition was superheated. These lines were all straight with slopes very similar to those obtained with supersaturated steam. Curves C and D are for wet steam. The former is for steam so wet that the temperature of the wire barely increased. For the other, in which the wetness was

less, the moisture was soon evaporated, when the curve turned over to become approximately parallel to the straight lines representing superheated and supersaturated steam.
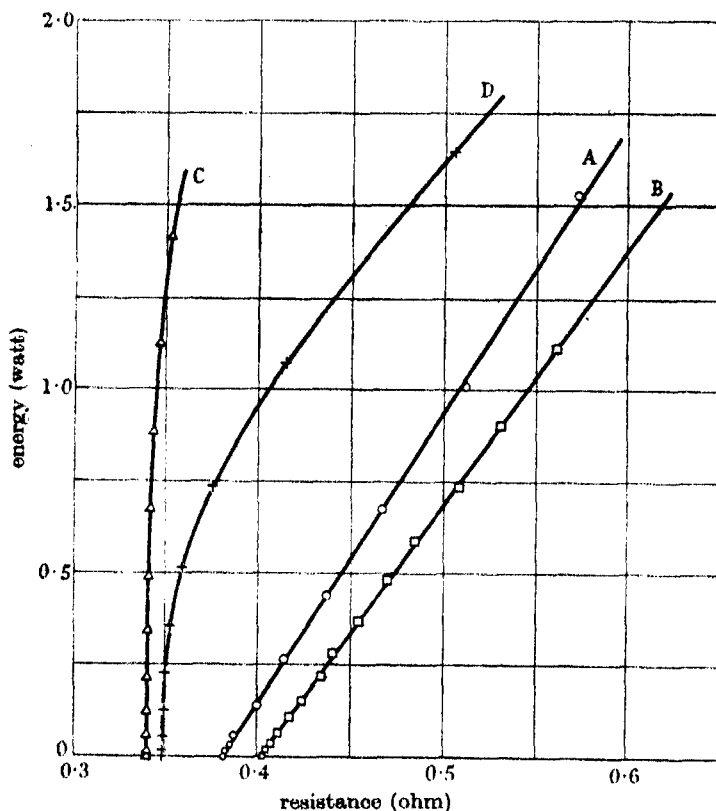
FIGURE 12. Nusselt numbers for superheated, supersaturated and wet steam. A, test no. 133b, supersaturated; B, test no. 128, superheated; C, test no. 132, wetness 6·7 %; D, test no. 133c, wetness 3·7 %.

6. DETERMINATION OF THE LOSSES IN THE NOZZLE AND THE POSITION OF THE WILSON LINE

For use in the preceding sections it was necessary to calculate from the observed pressure distributions the velocities and temperatures in the supersaturated region in the nozzle. As a first step the loss of heat drop had to be determined. This quantity, while called for convenience a friction loss, includes not only the friction loss but also the discrepancies arising from the Reynolds theory, which assumes the flow to be one-dimensional and takes no account of the curvature of the stream lines. In the earlier work the loss was ascertained by a laborious trial and error method, but an improved procedure, which permits of direct calculation, has since been devised.

Consider the expansion from an initial pressure p_0 to a tapping, where the pressure given by the Reynolds theory on the assumption of isentropic expansion is p_1 and where the observed pressure is p_1'. Let $(H_0 - H_1)$ and $(H_0 - H_1')$ be the corresponding isentropic and actual heat drops. Then, if $(1 - K^2)$ is the fractional loss of heat drop,

$$(H_0 - H_1') = K^2(H_0 - H_1). \tag{6.1}$$

It is shown in the Appendix that K may be calculated from the known values of p_0, p_1 and p_1' by means of the equation

$$K^2\{(p_0/p_1)^{3/13} - 1\} + Kp_1'/p_1 - (p_0/p_1)^{3/13} = 0. \tag{6.2}$$

The numerous observations of p_1'/p_0 at each tapping were meaned, and were inserted in turn into (6.2), the necessary properties of steam being taken from the tables of Callendar & Egerton (1939). The results are shown in figure 13 under the title nozzle 2, and a mean straight line has been drawn through the plotted points. Two departures from this line are apparent. Near the throat a hump occurs, which can be attributed to the curvature of the stream-lines in that region, a factor ignored
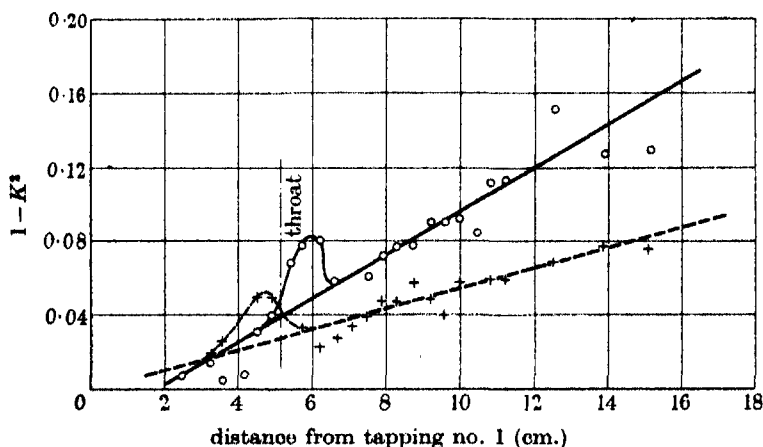
FIGURE 13. Loss of heat drop due to friction. ○ nozzle no. 1; + nozzle no. 2.

by the Reynolds theory. Farther downstream there are traces of a wave, which was noticed when the mean values of p_1'/p_0 were plotted; a similar wave had been observed in the previous experiments. From the mean straight line the velocities and temperatures were deduced, which were required in the preceding sections. The existence of the hump was not important because no results in that region were wanted. For purposes of comparison the results of the earlier work were recalculated with the new steam tables and are plotted on the same figure under the title nozzle 1. They are of the same general form, but the loss was greater, this being no doubt due to the smaller hydraulic mean depth and to the greater angle of divergence. In order to

make possible the prediction of the pressure at a given tapping in another nozzle operating under specified entry conditions, the analysis was extended to include the gain of entropy during the expansion. It is proved in the Appendix that this entropy increase

$$\phi_1' - \phi_1 = 1\cdot09876 \log K + 0\cdot084520 \log p_1'/p_1. \qquad (6\cdot3)$$

This expression was evaluated for the two nozzles, and the results are plotted in figure 14, where the hump again appears but in an inverted form. If in the design stage it is desired to estimate the pressure at a tapping in a new nozzle, the probable values of K and $(\phi' - \phi)$ can be obtained from figures 13 and 14, and the corresponding point plotted on the H-ϕ chart.
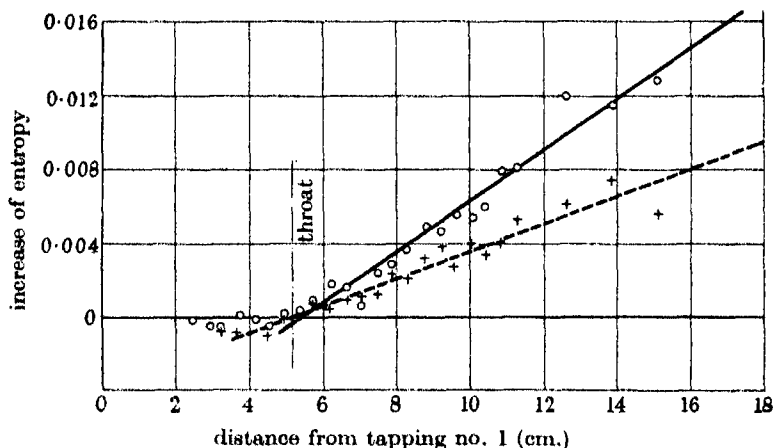
FIGURE 14. Increase of entropy due to friction. ○ nozzle no. 1; + nozzle no. 2.

To show the Wilson line on the H-ϕ chart the pressures at the beginning of condensation were read for thirty-one tests divided into three sets, each at approximately constant inlet pressure, and the necessary friction corrections were applied as explained above. For purposes of comparison the observations with nozzle 1 were recalculated with the new steam tables and the new method for finding the friction correction. The results are to be found in figure 15, where Fahrenheit units have been reverted to in order that comparison may more easily be made with published charts. Three points are shown for each test, corresponding to the steam conditions A at inlet, B after isentropic expansion to the pressure observed at the beginning of condensation, and C after the friction correction had been applied to B. For each test the points A and B are shown by the same symbol, to which a circle has been added to indicate C. The points C form the Wilson line, the position of which is seen to depend as much on the initial steam conditions as on the shape of the nozzle. But it is worthy of note that the points B all lie in a very narrow band.

This result may be due to friction at the walls of the nozzle, which violates the assumption (made in the calculations) of uniformity of conditions over cross-sections.
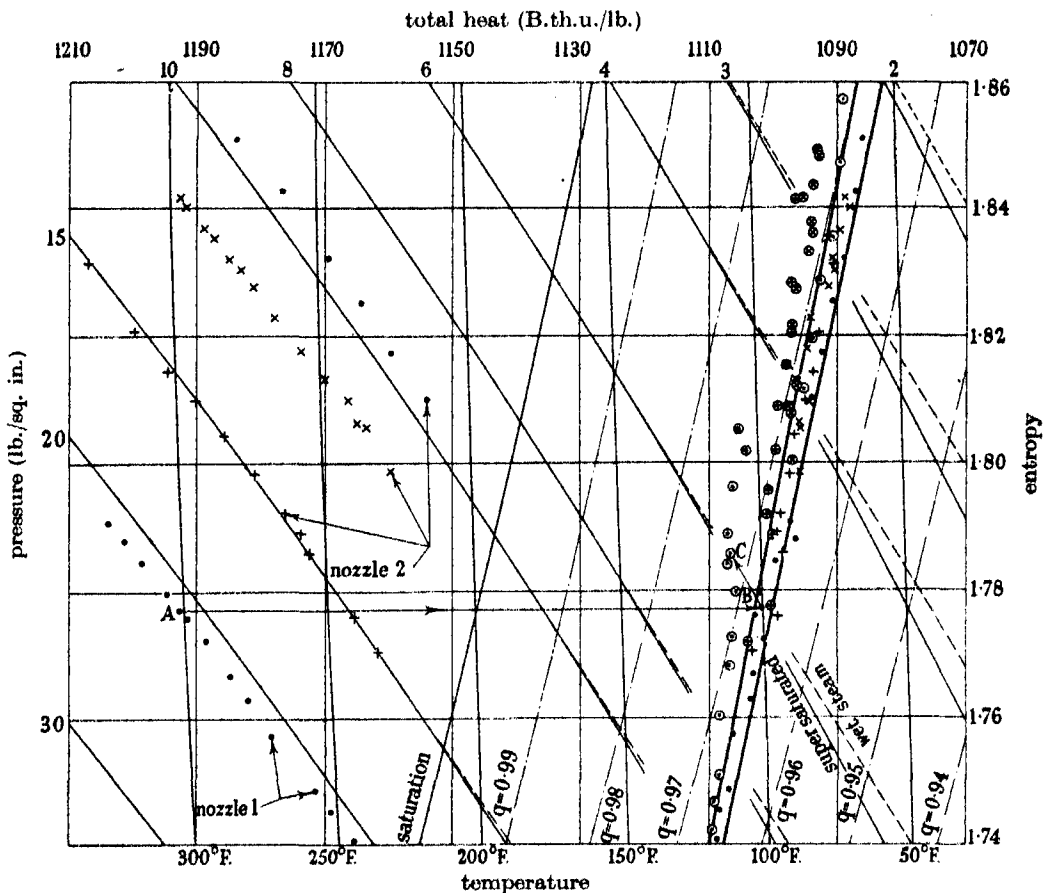
FIGURE 15. The Wilson line.

7. CONCLUSIONS

(1) The work described in the preceding sections showed that, without the use of very refined methods of resistance measurement, the wire detected the appearance of moisture brought about by the change from the supersaturated to the wet state. The wire also indicated the removal of moisture during the process of recompression, and (as expected) it proved insensitive to the difference between superheated and supersaturated steam. No examination was made of the change direct from the superheated to the wet state, whether caused by loss of heat to the surroundings or by slowly doing work. Our intention of carrying out experiments on this subject (together with further research on different wires) was frustrated by the outbreak

of war, but it seems improbable that the results would have been inconsistent with those previously obtained. It appears that the presence of moisture can be easily detected by determining the resistance-current relation for the wire. If the steam is not wet, the resistance (and hence the temperature) of the wire immediately begins to rise as the current is increased; moisture in the steam shows itself by markedly delaying the rise in the resistance.

(2) The temperature taken up by the unheated wire was found to be greatly in excess of that of the supersaturated steam in which it was immersed. As predicted by Griffith's theory, the wire temperature was nearly the same as the steam temperature at entrance to the nozzle. The small excess of observed over theoretical temperature was attributed to the presence of shock waves.

(3) Over the small range of Mach numbers (exceeding unity) and of Reynolds numbers which could be investigated, the Nusselt number of heat transfer from the heated wire to the surrounding steam was about 15.

(4) The position of the Wilson line was compared with that determined for the different nozzle used in the earlier work: it depended as much on the steam conditions at inlet as on the shape of the nozzle.

APPENDIX. DETERMINATION OF THE FRICTION LOSS IN THE NOZZLE

Throughout the main part of this paper use was made (except where otherwise stated) of the steam tables of Callendar & Egerton (1939) and of charts plotted from them. But for the purpose of determining the friction loss by the new method it was necessary to employ characteristic equations for steam, and these are not given in the tables mentioned above. An approximate method of determining the correction was, however, legitimate, and the equation

$$H = p(V - 0.0123)/2.2436 + 464,\qquad\text{(A. 1)}$$

given by Callendar (1930) for superheated and supersaturated steam was employed. H is the total heat per lb. in lb. C.H.U., p the pressure in lb./sq. in., and V the specific volume in cu.ft./lb. The second term inside the bracket may here be ignored because the experiments were carried out with steam of large specific volume. As in §6 let the suffixes 0 and 1 refer to isentropic expansion from the inlet to the tapping, with primes to indicate the actual conditions obtaining at that tapping. In addition, suppose A to be the sectional area of the nozzle, and the suffix x to refer to a cross-

section near the inlet where the isentropic and observed pressures are indistinguishable. Now from (6·1) the mean velocities v are related by

$$v_1'/v_1 = \{(H_0 - H_1')/(H_0 - H_1)\}^{\frac{1}{2}} = K, \tag{A.2}$$

and the mass discharge m under both isentropic and actual conditions is

$$m = A_x v_x/V_x = A_1 v_1/V_1 = A_1 v_1'/V_1', \tag{A.3}$$

so that
$$V_1'/V_1 = K. \tag{A.4}$$

From (A.1)
$$H_0 - H_1 = (p_0 V_0 - p_1 V_1)/2 \cdot 2436, \;\Big\}$$

and
$$H_0 - H_1' = (p_0 V_0 - p_1' V_1')/2 \cdot 2436; \Big\} \tag{A.5}$$

and on substituting (A.5) and (A.4) in (6·1) we obtain

$$K^2(p_0 V_0/p_1 V_1 - 1) + K p_1'/p_1 - p_0 V_0/p_1 V_1 = 0. \tag{A.6}$$

But the index of isentropic expansion is $1 \cdot 30$ and $p_0 V_0/p_1 V_1 = (p_0/p_1)^{3/13}$, hence (A.6) becomes

$$K^2\{(p_0/p_1)^{3/13} - 1\} + K p_1'/p_1 - (p_0/p_1)^{3/13} = 0, \tag{6·2 bis}$$

which is the required result.

To obtain the corresponding increase of entropy ϕ we employ the Callendar equation

$$\phi = 1 \cdot 09876 \log (T/373 \cdot 1) - 0 \cdot 25356 \log (p/14 \cdot 689) - 2 \cdot 4cp/7T + 1 \cdot 76300, \tag{A.7}$$

where T is the absolute temperature in degrees centigrade, and

$$c = 0 \cdot 4213(373 \cdot 1 T)^{10/3}.$$

Then the required increase of entropy is

$$\phi_1' - \phi_1 = 1 \cdot 09876 \log (T_1'/T_1) - 0 \cdot 25356 \log (p_1'/p) - 2 \cdot 4(c_1' p_1'/T_1' - c_1 p_1/T_1)/7. \tag{A.8}$$

Callendar's equation of state is

$$V = 1 \cdot 0706 T/p - c + 0 \cdot 01602, \tag{A.9}$$

in which the last two terms are so small in comparison with V for low-pressure steam that they may be neglected. Hence from (A.9) and (A.4)

$$T_1'/T_1 = p_1' V_1'/p_1 V_1 = K p_1'/p_1. \tag{A.10}$$

Then with the aid of (A.9) and (A.10), (A.8) becomes

$$\phi_1' - \phi_1 = 1 \cdot 09876 \log K + 0 \cdot 84520 \log (p_1'/p_1) - 0 \cdot 3671(c_1' - c_1 k)/V_1'. \tag{A.11}$$

A representative value of the last term is 3 % of $(\phi_1' - \phi_1)$, and this may be ignored. Thus (A.11) reduces to

$$\phi_1' - \phi_1 = 1 \cdot 09876 \log K + 0 \cdot 84520 \log p_1'/p_1, \tag{6·3 bis}$$

as stated in § 6.

REFERENCES

Binnie & Woods 1938 *Proc. Instn Mech. Engrs,* 138, 229.
Callendar 1920 *The properties of steam.* London: Arnold.
Callendar 1930 *Abridged Callendar steam tables, centigrade units,* 2nd ed. London: Arnold.
Callendar & Egerton 1939 *The 1939 Callendar steam tables.* London: Arnold.
Goldstein (ed.). 1938 *Modern developments in fluid dynamics,* p. 607. Oxford: Clarendon Press.
Griffith (unpublished); see Goldstein, p. 631.
Hawkins, Solberg & Potter 1935 *Trans. Amer. Soc. Mech. Engrs,* p. 395.
Hilpert 1932 *Forschungsh. Ver. dtsch. Ing.* p. 355.
Hilpert 1933 *Forsch. Arb. IngWes.* 4, 215; see Goldstein, p. 636.
Hilton (unpublished); see Goldstein, p. 631.
Hilton 1938 *Proc. Roy. Soc.* A, 168, 43.
Martin 1918 *Engineering,* 106, 1.
Meissner 1938 *Forsch. Arb. IngWes.* 9, 213.
Milverton 1935 *Proc. Roy. Soc.* A, 150, 287.
Müller 1920 *Z. ges. Turbinenw.* 17, 61.

Precipitation in single crystals of silver-rich and copper-rich alloys of the silver-copper system

By F. W. Jones,* Ph.D., P. Leech,† M.Sc. and C. Sykes,* D.Sc.

(Plates 4, 5)

The course of precipitation at different temperatures in silver-rich and copper-rich single crystals of silver-copper alloys of various compositions has been studied by X-ray methods. Precipitation is continuous at the higher temperatures when the degree of supersaturation is low and discontinuous when the temperature is low and the supersaturation is high. The experimental results are considered in relation to the various theories of the precipitation process.

1. Introduction

X-ray photographs of many age-hardening alloys taken during the precipitation process show that the change in composition of the solid solution is continuous, i.e. there is a gradual movement of the X-ray lines of the solid solution, from positions corresponding to the lattice spacing of the original composition to new positions corresponding to the spacing of the equilibrium composition. This has been called 'continuous precipitation'. On the other hand, several investigators (Ageew, Hanson & Sachs 1930; Cohen 1937) have established that in polycrystalline silver-copper alloys 'discontinuous precipitation' occurs. Throughout the

* National Physical Laboratory.
† Messrs Metropolitan-Vickers Research Department.

Jones, Leech and Sykes

Proc. Roy. Soc., A, volume 181, plate 4
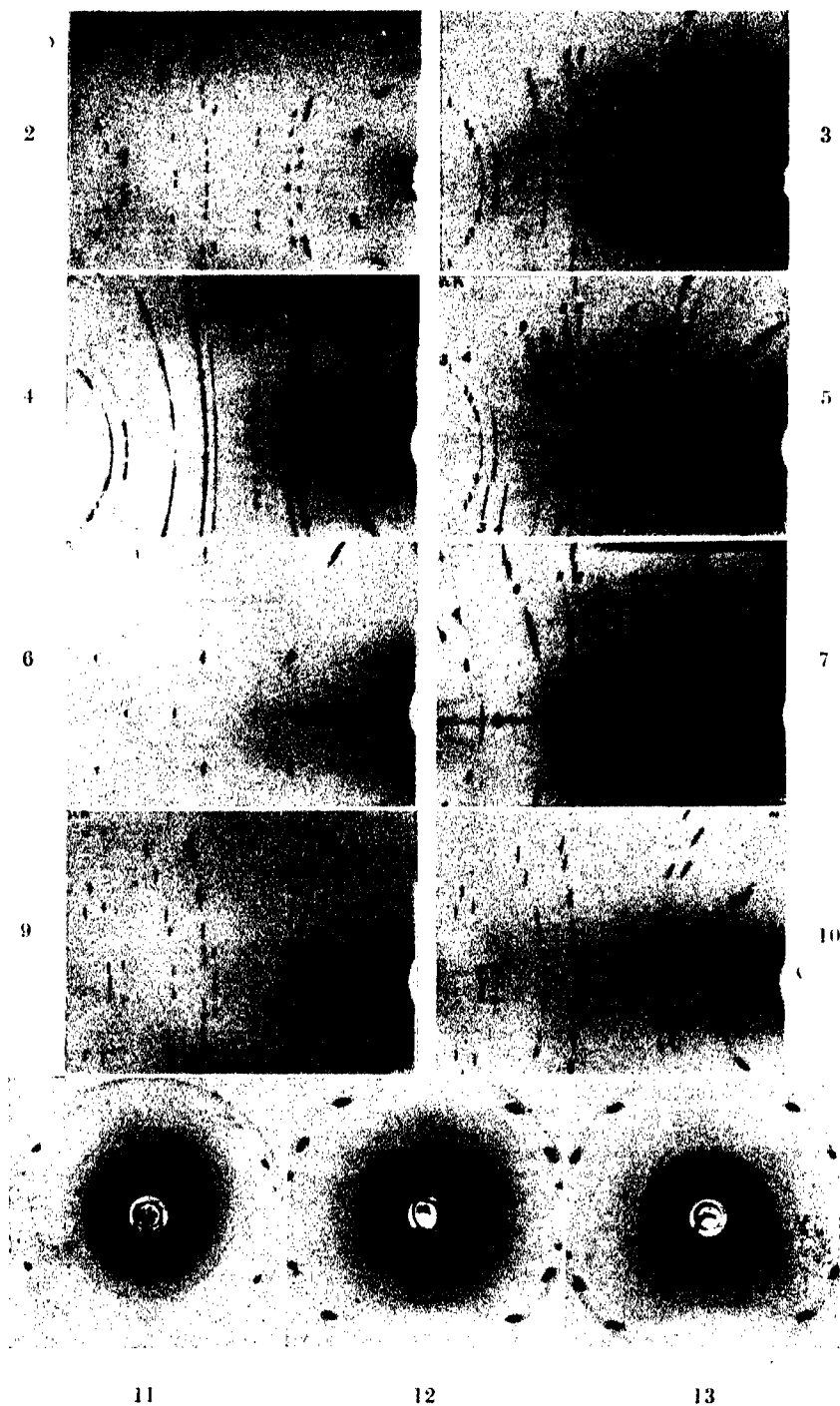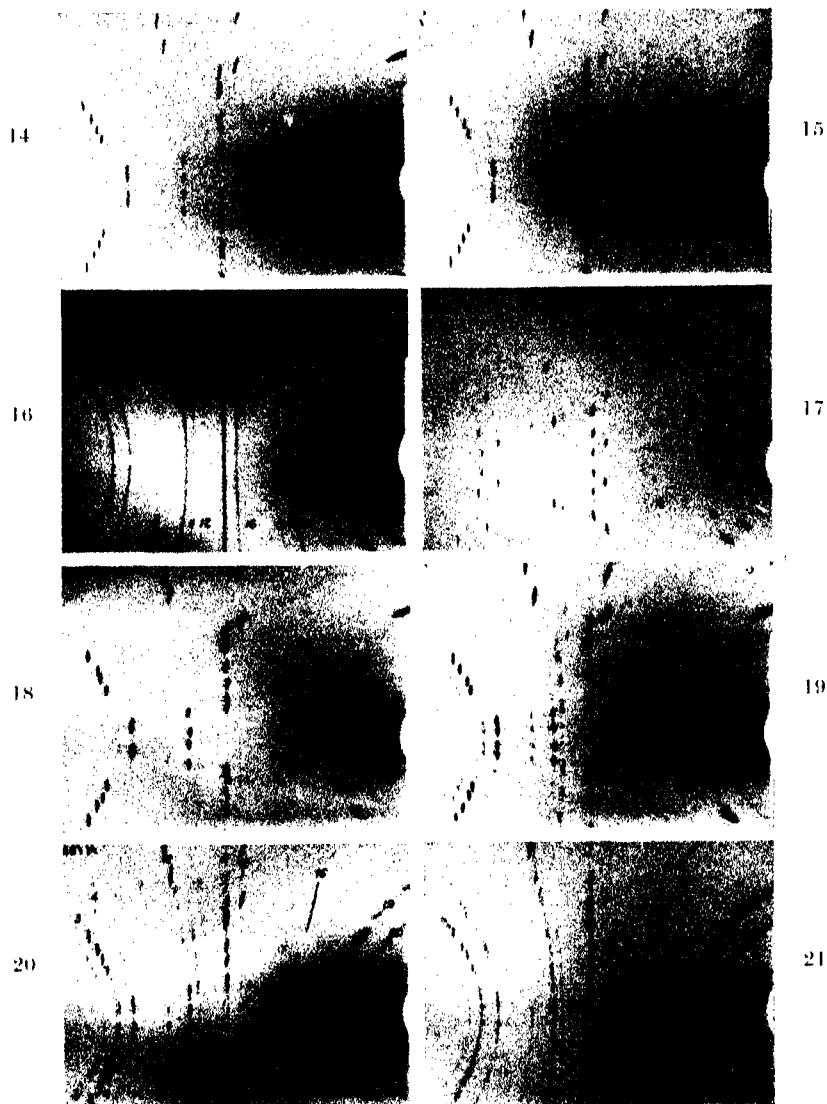
ageing process the positions of the X-ray lines corresponding to the original super-saturated solid solution remain unchanged, but the lines decrease in intensity, whilst a new set of lines appears corresponding to the spacing of the stable solid solution, and these increase in intensity. This suggests that the change in composition of the solid solution is discontinuous, a result which is somewhat difficult to explain. One possible interpretation is that an alloy in an intermediate stage in the precipitation process consists partly of crystals in which precipitation is complete and partly of crystals in which precipitation has not commenced. This hypothesis can be tested by experiments on single crystals.

Work with single crystals also makes possible the study of orientation relationships, recrystallization phenomena and other effects such as those observed by Preston (1938) with aluminium-copper alloys. It cannot be tacitly assumed, however, that results obtained on single crystals also hold good for polycrystalline specimens, since several instances have been reported of differences in behaviour during precipitation (Schmid & Siebel 1934). In spite of this, the more detailed information obtainable from X-ray photographs of single crystals suggests that the results may be of considerable interest in studying the mechanism of the precipitation process.

Ageew *et al.* (1930) have published back-reflexion photographs of a single crystal of a silver-copper alloy containing 5 % copper which showed that recrystallization occurred as a consequence of the strains set up by precipitation. Wiest (1932, 1933) has also taken photographs of single crystals of copper-silver alloys containing 5, 7 and 10 % of silver, but apart from stating that precipitation was continuous, unlike the discontinuous precipitation observed with polycrystalline material, he gives few details of the effects observed in his photographs.

The present paper records the results of an X-ray examination of single crystals of silver-copper and copper-silver alloys during ageing at various temperatures.

2. Experimental

A. *Materials*

The alloys of which it was desired to grow single crystals were obtained in the form of wire about 1 mm. diameter. They were prepared from high-purity constituents. The silver contained 99·9 % silver by weight and the copper 99·98 %. The crystals as grown were about 1·25 mm. diameter and 10 cm. long. For use as X-ray specimens they were cut into pieces about 12 mm. long, a length of about 1 mm. at the centre of the piece being examined by X-rays. During the production of the single crystals some segregation occurred, so that the composition of any particular piece was not necessarily that of the original material. The compositions of the individual pieces were determined by measuring the lattice parameter after a long solution heat treatment and using the lattice parameter-composition curves previously determined by X-ray workers (Ageew, Hanson & Sachs 1930; Ageew & Sachs 1930; Schmid & Siebel 1933; Owen & Rogers 1935).

B. *Growth of the single crystals*

The single crystals were prepared by the Bridgman method by lowering through a furnace having a region 50–100° C above the melting-point. The apparatus is shown diagrammatically in figure 1. The crucible (*G*) is a 6 mm. diameter rod of Acheson graphite having an axial hole 1·25 mm. diameter (closed at the bottom end) down which the alloy wire is pushed. The crucible is suspended on a 0·1 mm. diameter tungsten wire which is wound round a drum (*E*), the drum being rotated by an electric clock mechanism inside the vacuum chamber (*F*). The specimen is lowered through the silica tube (*S*) at about 10 cm. an hour. At (*L*), (*M*) and (*N*) are ground-joints sealed with Apiezon Sealing Compound *Q*. The system is evacuated by an oil diffusion pump and rotary backing pump.

At first attempts were made to carry out the solution heat treatment in the same furnace, immediately after solidification of the crystal. The furnace was wound to have a region *CD* at the solution temperature, just below the region *AB* in which the alloy had been melted. After cooling through the region *BC* the specimen was held in the region *CD* for several hours and then crucible and specimen were quenched. This method was not satisfactory, as the presence of the crucible gave rise to a quenching rate which was too slow for the silver-rich alloys and precipitation occurred. Furthermore, if the badly quenched crystal was subsequently reheated for requenching, the crystal broke up.

FIGURE 1

The following method was found to be successful and was adopted. The specimen was lowered in the crucible right through the furnace. This produced a 'crystal' in which some precipitation had taken place, but the orientation of the matrix was the same throughout and the specimen could be reheated without disturbing the continuity of the crystal.

The specimen so prepared was carefully removed from the graphite crucible and heat-treated as described in § C.

C. *Heat treatment*

High-angle X-ray spots on back-reflexion photographs of crystals quenched after heat treatment for 10 hr. in vacuo at 780° C were not very sharp. Crystals were then annealed for 100 hr. at 780° C in hydrogen (to prevent oxidation and preferential evaporation of one constituent); the hydrogen was then removed by
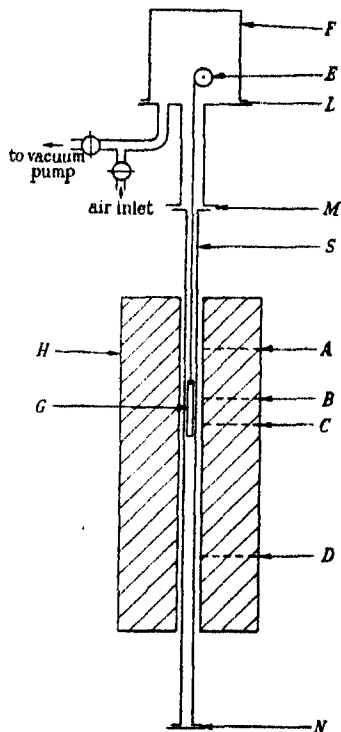
evacuation and the specimen quenched by the admission of water into the evacuated furnace tube (Rosenhain 1908). This treatment gave satisfactory photographs, and according to Schmid & Siebel (1933) should be more than sufficient to ensure that the alloy is a homogeneous solid solution.

Ageing was carried out in tubular resistance furnaces, the furnace temperature being maintained constant to $\pm \frac{1}{2}°$ C by an automatic temperature regulator of the platinum-resistance type. During ageing, specimens were sealed into evacuated pyrex tubes. Where ageing times were less than 10 min. the pyrex tube containing the specimen was immersed in a bath of molten tin.

D. *X-ray apparatus*

The X-ray rotation photographs were taken in a cylindrical camera 3 in. in diameter, with the beam glancing one side of the specimen, which was wider than the beam. Usually the axis of rotation was the axis of the cylindrical crystal rather than any particular crystallographic direction, but some rotation and oscillation photographs have been taken with particular crystallographic directions as axes. In some cases lattice spacing measurements were made on the crystals by the back-reflexion method. The distance between specimen and film was found from the position of the lines of a calibration substance (usually silver, of which a small quantity, in the form of fine powder, was coated on the surface of the single crystal).

All rotation photographs have been taken with copper $K\alpha$ radiation; for some of the back-reflexion photographs cobalt or nickel radiation was used. All photographs were taken with the specimen at room temperature ($19 \pm 2°$ C).

3. Experimental results

A. *Silver-rich alloys*

(a) *Ageing at* 300° C. Figures 2–5 (plate 4) illustrate the changes which occur during ageing of a single crystal containing 7·3 % copper by weight. Reflexions from one side of the specimen only appear on the photographs. θ, the Bragg angle, increases from left to right. The spots may be identified by means of figure 5, on which the values of $(H^2 + K^2 + L^2)$ are given for the solid solution (at the top) and the precipitate (at the bottom). After 8 min. (figure 3) some precipitation has taken place. Some parts of the single crystal have retained the original orientation and lattice spacing, whilst other regions have broken up into material of somewhat different orientation and a larger lattice spacing than that of the single crystal. As far as can be ascertained from measurement of the films, this material is the solid solution composition stable at 300° C. From the width of the lines of the stable composition it may be estimated that these regions are not smaller than 150 A, i.e. about 40 atomic distances linear dimensions. The high-angle spots on the photograph show smears on one side, indicating the presence

of some material having lattice spacings intermediate between those of the unstable and stable solid solutions. If this effect was present on photographs of polycrystalline specimens, it is unlikely that it could be observed with any certainty. It may be due to strain—either tensile strain in the unstable or compressive strain in the stable solid solution—but as the smears appear to lie only between the lines of the stable and unstable solid solutions, we think a more likely explanation is that there is some material present which has compositions intermediate between those of the stable and unstable solid solutions. If this is so, besides those regions in which precipitation has not commenced and those in which precipitation is complete, there are some regions in which precipitation is in an intermediate stage. This material has the same orientation as the original single crystal, so it seems that a region of the crystal retains its orientation, the matrix composition changing gradually, until it approaches the stable solid solution composition, i.e. precipitation is almost complete in the region when crystal fragmentation* takes place. It appears possible, therefore, that even in this alloy true 'discontinuous precipitation' does not take place, but there is at any one time only a small amount of material in the transition stage.

After 90 min. (figure 4) precipitation is nearly complete. All the original single crystal spots have disappeared. It is noticeable that there is some preferential orientation in the stable solid solution. There is as yet no trace of the precipitated copper. The precipitate lines can, however, be seen long before precipitation is complete if monochromatic radiation is used (Cox & Sykes 1940).

Further, long annealing at the ageing temperature results in the matrix slowly returning to its original orientation (compare figures 2–5). Figure 5 shows a photograph of the crystal after 1 month at 300° C. On the original photograph lines of the precipitated copper can be readily seen, and these show that the orientation of the precipitate was random. If there were any preferred orientation of the precipitate, it should be more easily detected on a photograph taken with the crystal rotating about the axis (001), or some other crystallographic direction of low Miller indices. A photograph was therefore taken of the crystal after heat treatment for 1 month at 300° C, the specimen being rotated about the (001) axis. The photograph showed continuous lines of the copper precipitate; the intensities of the lines appeared to be constant along their length, so that the preferential orientation of the precipitate, if any, is very slight.

Results with a crystal containing 4·7 % copper were very similar to those obtained with the crystal containing 7·3 % copper. The precipitation process in the 7·3 % copper alloy was about four times faster than in the 4·7 % copper alloy. The original orientation of the 4·7 % copper crystal was the same as that of the 7·3 % copper crystal. Comparison of the photographs showed that the type of preferential orientation of the matrix in the two crystals was the same after precipitation. For the 4·7 % copper crystal this orientation was determined as follows. The crystal was orientated so that the axis of rotation was perpendicular to the

* Referred to subsequently as 'fragmentation'.

(001) plane and the position of the (100) and (010) normals determined by two 10° oscillation photographs. The crystal was then aged for 4 hr. at 300° C and replaced in the goniometer so that its orientation was the same as before ageing. Figures 6 and 7 (plate 4) show rotation photographs of the crystal before and after ageing. A series of 10° oscillation photographs was then taken and pole figures were plotted for indices (100), (110), (111) and (311). The pole figure found for (100) is shown in figure 8. All pole figures correspond to that obtained by giving the crystal rotations of between 35 and 55° about the normals to the three cube faces as axes. This is in good agreement with the result obtained by Barrett, Kaiser & Mehl (1935), i.e. a rotation of $42 \pm 5°$ about the same axes.
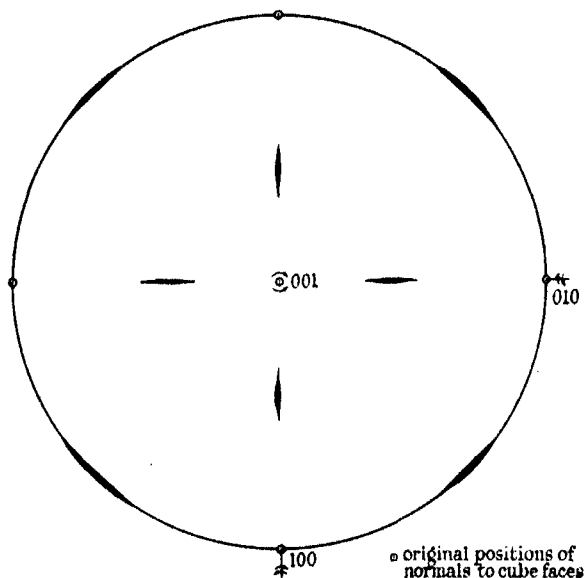
FIGURE 8

We can suggest no hypothesis which would account for this rotation.

(b) *Ageing at* 550° C. Practically no breaking up of the single crystal takes place when a 4·7 % copper alloy is aged at 550° C. Figures 9 and 10 (plate 4) show photographs of the crystal as quenched and after precipitation is complete (20 min. at 550° C). Figure 10 shows one or two extra spots, due to some of the matrix having a different orientation, whilst there are also faint spots in positions corresponding to a precipitate of copper, having the same orientation as the matrix. They are too faint to be reproducible on figure 10, but can be clearly seen on the original photograph.

The change in spacing appears to be continuous at 550° C. This is best illustrated by the back-reflexion photographs shown in figures 11–13 (plate 4). Figure 11 is the crystal as quenched, figure 12 after 5 min. at 550° C and figure 13 after 20 min. The continuous lines are due to pure silver, used as a calibration substance.

(c) *Precipitation during cooling.* In experiments in which precipitation occurred during cooling, the orientation of the matrix always remained the same, i.e. there was no fragmentation at any of the cooling rates used (quenching to cooling at 30° C/hr.). At a rate of 30° C/hr. the precipitate showed some preference for the orientation of the matrix, but it was estimated that more than half of it was randomly oriented. The ageing experiments at 300 and 550° C suggest that the orientation of the material precipitated at low temperatures is random, at high temperatures the same as that of the matrix. Thus the precipitate obtained on cooling is to be expected to be partly that of the matrix, partly random. This is in agreement with the work of Barrett *et al.* (1935). They found that the necessary conditions for development of Widmanstätten figures were slow cooling at high temperatures, followed by rapid cooling through lower temperatures. For the formation of Widmanstätten figures it is necessary that the orientation of the matrix and precipitate should be related.

B. *Copper-rich alloys*

(a) *Precipitation at* 380 *and* 500° C. The results obtained with copper-rich alloys vary with the composition of the alloy and the ageing temperature. At 380° C the ageing of a 5·3 % silver alloy proceeds similarly to silver-rich alloys aged at 300° C. Figures 14–16 (plate 5), as quenched, and aged 18 hr. and 6 weeks respectively, illustrate the changes which take place in the alloy during precipitation. The precipitation is largely discontinuous and is accompanied by fragmentation. As long as the original X-ray spots of the solid solution are sharp enough to measure, the spacing is the same as that of the quenched alloy, within the limits of experimental error. Unlike the case of the silver-rich alloys, the orientation of the matrix after fragmentation is random, the crystal size being large enough to produce 'spotty' X-ray lines, which are also quite sharp, indicating that the fragmentation has relieved much of the strain caused by precipitation (see figure 16). As soon as the lines of the stable matrix are strong enough to measure, the spacing is that of the solid solution in equilibrium at the ageing temperature. As with the silver-rich alloys, during precipitation there appears to be a small amount of material present which has lattice spacings and hence compositions intermediate between those of the original and the stable solid solutions. This was shown by back-reflexion photographs taken with nickel radiation, which gave very high resolution.

The precipitate can be seen quite early in the precipitation process. The lines are just visible after 18 hr. at 380° C, although precipitation takes about 70 hr. to complete. Maximum hardness in polycrystalline specimens of these alloys is attained at about the time precipitation reaches completion (Cox & Sykes 1940). After 50 hr. ageing at 380° C the average size of the precipitate crystals, estimated from the breadth of the X-ray lines, was about 200 A. This method of estimation gives a lower limit for the estimation of crystal size, since broadening due to strain and composition variations is neglected. The fact that the lines are somewhat spotty suggests that the crystals are actually very much larger than this.

When a crystal containing 5·2 % silver is aged at 500° C the spacing (and hence the composition of the matrix) of the single crystal changes continuously, as table 1 shows. After 8 min. at 500° C, very faint spots due to precipitated silver can be seen on the photographs, the precipitate having the same orientation as the matrix. As precipitation proceeds these spots slowly increase in intensity (see figure 18, aged 20 min. and figure 19, aged 2 hr.). When the matrix fragments at the end of the precipitation process, the precipitate follows suit. The close relationship which is retained between the orientations of the matrix and the precipitate during fragmentation is very striking. When there is a particularly intense spot on one of the matrix lines a spot is also observed on the precipitate line of the same Miller indices, corresponding to the same crystal orientation. This is shown in figure 20 (aged 36 hr.) and figure 21 (aged 15 days).

TABLE 1

time at 500° C	0	$\frac{1}{2}$ hr.	1 hr.	2 hr.	6 hr.	equilibrium
spacing (A)	$3 \cdot 627_0$	$3 \cdot 624_4$	$3 \cdot 619_8$	$3 \cdot 615_0$	$3 \cdot 613_2$	$3 \cdot 613_1$

From the width of the precipitate spots, it may be estimated that 100 A is a lower limit to the average size of the precipitate crystals after $\frac{1}{2}$ hr. at 500° C. After 1 hr., when precipitation is about half complete (according to the lattice constant measurements), the lower limit is about 200 A. It is about the same when precipitation is nearly complete, after 2 hr. These figures are in reasonable agreement with the size given by Cox & Sykes (1940) for polycrystalline material. From the 'spottiness' of the precipitate lines after fragmentation, it appears that the size of some of the precipitate crystals must be of the order of 10^{-3} cm. linear dimensions. This gives an estimate of the distance which some of the silver atoms must travel before precipitation is complete.

Even at 500° C there is evidence that considerable strains are present during the precipitation process. That the silver-rich precipitate is under compressive strain is illustrated by comparison of the distances between the silver line $H^2 + K^2 + L^2 = 27$ and the knife edge on figures 19 and 21. In figure 21, the $K\alpha$ doublet is resolved indicating that strains have been removed by the long anneal at 500° C. Back-reflexion photographs have been taken using the specimens treated for 2 hr. at 500° C and 360 hr. at 500° C, and the lattice spacing of the precipitate determined. The difference in spacing is about 1 part in 200, corresponding to a hydrostatic pressure of 100 tons/sq. in.

In this crystal it seems that the strains become so large when precipitation is complete that the matrix collapses and fragmentation takes place.

Unlike the silver-rich alloys, the copper-rich alloys showed no tendency to return to the same orientation as the original single crystal after long ageing.

Another crystal containing 4·0 % silver precipitated continuously when aged at 380° C. The photographs were similar to figures 18–21 (plate 5), but fragmentation commenced well before precipitation was complete. At 500° C precipitation was continuous, with no fragmentation, in a crystal containing 4·1 % silver. The results

at 380 and 500° C are summarized in table 2 below. It will be noted from this table that fragmentation may occur when the precipitation is continuous.

TABLE 2

wt. % of Ag	temp. °C	precipitation	
5·3	380	discontinuous	fragmentation
5·2	500	continuous	,,
4·0	380	continuous	,,
4·1	500	continuous	no fragmentation

(b) *Precipitation on cooling*. Like the silver-rich crystals, crystals of copper-rich alloys showed no fragmentation when precipitation occurred during cooling at a constant rate. The amount of randomly orientated precipitate was, however, much smaller with copper-rich alloys than silver-rich alloys (see figure 17, plate 5, which shows a copper-silver alloy containing 5·2 % silver cooled from 780 at 30° C/hr.). Much of the precipitate has the same orientation as the matrix.

C. *Sundry observations on silver-rich and copper-rich crystals*

(a) *Pre-precipitation phenomena*. In an aluminium-copper alloy containing 4 % copper, Preston (1938) has shown by means of Laue photographs of single crystals that the initial stage of precipitation is the formation of plates of copper-rich atoms on the solid solution lattice. Barrett & Geisler (1940) have reported similar effects in an aluminium-silver alloy containing 20 % silver.

A series of Laue photographs was taken of a single crystal of a copper-silver alloy containing 4 % silver, aged for various periods of time at 400° C. No changes in the photographs were observed which might have been attributed to plate formation in the early stages of precipitation. The photographs showed some diffuse streaks which were found to be identical with those given by a single crystal of pure copper, and these streaks remained unchanged throughout the course of precipitation. Such streaks have been the subject of several papers in *Proc. Roy. Soc.* (1941).

Precipitation was continuous, and the crystal broke up as precipitation neared completion.

(b) *Influence of quenching rate*. Copper-rich alloys are not particularly sensitive to variations in the quenching rate, but silver-rich alloys require fast quenching. Crystals of silver-rich alloys quenched in carbon crucibles showed smears on one side of the high-angle X-ray lines (even after long solution treatments at 780° C), indicating the presence of material having a lower copper content than the rest of the crystal. One crystal quenched in air showed two sets of spots having different spacings. The lack of reproducibility of the hardness-time curves of these alloys observed by Ageew *et al.* (1930) is probably to be attributed to the sensitivity of these alloys to variations in the quenching rate. Sykes, Swindells & Blears (unpublished work) also found that the specific-heat temperature curve of a silver-

copper alloy containing $7\frac{1}{2}$ % copper, in the quenched condition, was not reproducible, and that if the quenching rate was reduced by oil quenching, the curve was considerably modified.

The Rosenhain method of quenching which was used in the experiments described usually gave sharp high-angle spots, free from smearing, on the X-ray photographs.

(c) *Break-up of the single crystals.* It appears that if breaking up of the single crystals is to be prevented, it is necessary to avoid much precipitation taking place at low temperatures. A silver-rich single crystal which had been quenched broke up on reheating to 780° C in 1 hr., presumably due to precipitation having taken place on the way up. No break-up occurred when a quenched crystal was rapidly reheated to 780° C by immersing in a bath of molten tin. When crystals are cooled from 780° C to room temperature in a few hours, at a roughly uniform rate, it appears that most of the precipitation takes place at high temperatures, and no fragmentation occurs. Such crystals may be put into a furnace and reheated to 780° C without breaking up. Ageew *et al.* (1930) also found that a crystal which had been very slowly cooled broke up on rapid reheating to 780° C.

The copper-rich alloys are more sluggish than the silver-rich alloys. Quenched copper-rich crystals may be heated from room temperature to 780° C in an hour without fragmentation taking place. It is probable that at this heating rate there is not sufficient time for appreciable precipitation to occur at low temperatures.

(d) *Experiments with polycrystalline specimens.* We have attempted to determine by X-ray examination the type of precipitation occurring at relatively high temperatures in polycrystalline silver-rich and copper-rich alloys which had been cold worked and recrystallized. Precipitation was discontinuous in a silver-rich alloy, containing 5 % copper aged at 550° C, and in a copper-rich alloy containing 6 % silver aged at 600° C. At higher temperatures, or with smaller supersaturations, precipitation may be continuous, but the spacing change accompanying precipitation is so small that it is difficult to decide with any certainty the type of precipitation.

A copper-rich single crystal containing 4·1 % silver, in which precipitation was continuous at 500° C, with no break-up of the single crystal, was drawn down from 1·25 mm. diameter to 0·5 mm. The polycrystalline specimen obtained in this way was given a solution heat treatment and then aged at 500° C. Precipitation was discontinuous and about four times as rapid as in the single crystal.

Wiest & Dehlinger (1934) and Bumm & Dehlinger (1936) working with copper-silver alloys, and Schmid & Siebel (1934) (magnesium-aluminium alloys) have also found differences in behaviour between cast single crystals and material which has been cold worked and recrystallized. Jones & Leech (1941) found variations in the specific-heat temperature curve on copper-beryllium alloys according to the method of preparation.

4. Discussion of the Results

A. *General*

Probably the simplest theory of precipitation from a supersaturated solid solution is as follows:

Diffusion of the solute atoms occurs by interchange of atoms. At some places in the solid solution groups of atoms (or nuclei) will be formed, and these will grow as more precipitate atoms diffuse towards them through the solid solution, which will finally reach its equilibrium concentration.

Becker (1937, 1938, 1940) has attempted to calculate the effect on the course of the precipitation process of the surface energy of the nuclei which must be formed first. He found that such nuclei will be unstable and will redissolve unless they are larger than a certain minimum size depending on the degree of supersaturation of the solid solution. According to Becker the rate of nucleus formation is proportional to $\exp\left(-\dfrac{Q+A}{RT}\right)$, where Q is the diffusion activation energy and A is the free energy increase when a nucleus of the minimum stable size is formed. To a first approximation A is proportional to $1/(\Delta\alpha)^2$, where α is the solute concentration and $\Delta\alpha$ the degree of supersaturation. Thus, bearing in mind that in a given alloy the degree of supersaturation increases with a reduction in the ageing temperature, Becker's theory gives a qualitative explanation of the fact that at a given temperature the rate of precipitation increases with degree of supersaturation.

Once a nucleus of precipitate large enough to be stable has been formed, it will grow by taking up atoms from the surrounding matrix. Our conception of precipitation, following nucleation in an alloy for which Becker's assumptions hold, is then represented in figure 22 which shows the concentration (c) of precipitate atoms in the alloy as a function of the distance from the nucleus at A. B is the mid-point between two nuclei. (It is of course realized that 'concentration' has little meaning when applied to groups of only a few atoms.) The line (1) is at the concentration of the original supersaturated alloy. Curves (2), (3) and (4) represent successive stages of the precipitation process. The horizontal line (4) which represents complete precipitation is at the equilibrium concentration of the solid solution. When the nucleus is large enough to be stable the degree of supersaturation required to enable atoms
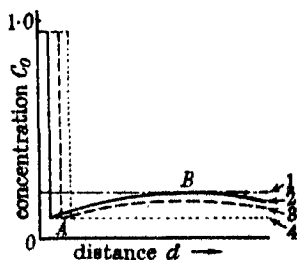
Figure 22

to cross the matrix-precipitate interface at A is probably quite small, so that the matrix concentration at the interface will be close to the equilibrium concentration. Besides 'negative diffusion' across the interface normal diffusion will proceed, caused by the concentration gradient between A and B bringing more precipitate atoms up to the interface. The shape of the concentration-distance curve will depend on the distances between nuclei and on the relative rates of negative and normal diffusion.

This description of precipitation agrees qualitatively with the results of the X-ray photographs of the copper-silver and silver-copper single crystals when precipitation takes place at high temperatures. The lattice spacing measurements indicate that the concentration of the solute decreases as precipitation proceeds.

On the other hand, Becker's treatment of the problem leads to no obvious explanation for the other type of precipitation—discontinuous precipitation—which is observed in the silver-copper system. In this connexion it may be noted that Becker assumed that the crystal structures and lattice parameters of the solid solution and precipitate were the same, and that the energy of the nuclei could be expressed solely in terms of the number of pairs of like and unlike atoms. In general the crystal structure and atomic volume of the precipitate will not both be the same as those of the matrix. Consequently there will usually be misfitting at the precipitate-matrix interface, and the surface energy will be similar to that of the usual grain boundary rather than the case considered by Becker. The energy of the nucleus should therefore include a further term to allow for the elastic energy due to the different atomic volumes of precipitate and matrix. It seems highly probable that this elastic energy term has a very important influence on the course of precipitation in many alloy systems. Mott & Nabarro (1940) and Nabarro (1940 *a, b*) have estimated the elastic energy of a nucleus of precipitate when the atomic volumes of precipitate and solvent are different, and their results will be considered in relation to the experimental results described in this paper.

B. *Orientation relationships and Widmanstätten patterns*

As mentioned above, Becker assumed that the crystal structures and lattice parameters of precipitate and matrix were the same. In actual cases where they are different, it is to be expected that the fit between the matrix and precipitate which has its equilibrium structure and spacing will be better for some planes of contact than for others, so that the energy of the precipitate-matrix interface will vary with the plane of contact. This assumption has been used by Mehl and his co-workers (1932, 1935, 1939) to provide a reasonable explanation for the Widmanstätten patterns which are observed in aluminium-copper and aluminium-silver alloys. They suggest that the precipitate takes the form of plates, suitably orientated in order to minimize the surface energy. At relatively high temperatures and with a low degree of supersaturation the results of Barrett *et al.* (1935) and the experiments with single crystals described in this paper show that in both silver-rich and copper-rich alloys of the silver-copper system, the orientations of the precipitate and matrix are the same. As the structures are the same the misfit will be equally bad on all planes of contact of the matrix-precipitate interface. Thus in this system it is difficult to see why the plates of precipitate should have any particularly favoured orientation. It is a fact that in alloys of the silver-copper system Widmanstätten patterns are much less readily developed than in the two aluminium alloys mentioned above. It is possible that the orientation of the

precipitate is not determined by considerations of good fit, since in all cases where the orientation of the precipitate has been investigated it either retains the same orientation as the solid solution, or where there is a change of structure, it takes up the orientation most closely related to that of the solid solution.

C. *The effect of different atomic sizes of solvent and solute: discontinuous and continuous precipitation*

In § 4 A we have mentioned various conditions which may govern the formation of nuclei of precipitate. In some cases stable nuclei are formed on the solid solution lattice, e.g. the copper-rich plates observed by Preston (1938) in aluminium-copper alloys. The resulting precipitate in this case has the composition $CuAl_2$ and so has a mean atomic volume much nearer that of aluminium than that of copper. Similar effects are not observed in the silver-copper alloys.

According to Nabarro (1940a) the atomic sizes of silver and copper are so different that it is unlikely that any stable nucleus can be formed and remain on the parent lattice, since its elastic energy under such circumstances would be greater than the chemical energy of precipitation; any nucleus will either break away immediately from the parent lattice or redissolve. It seems likely that the nuclei will form at faults or irregularities in the lattice, such as the boundaries between crystals or between the mosaic crystallites.

The subsequent growth of such nuclei has been considered by Nabarro (1940b). He first calculated the elastic energies of various shapes of precipitate and found that the plate-like forms have the lowest elastic energy. A plate which has broken away from the matrix will have a high surface energy, due to the misfitting of the precipitate and matrix at the common interface. The shape of the precipitate will therefore be a compromise between a thin plate of low elastic energy and a sphere of low surface energy. A further factor tends to thicken the plates. The number of atoms reaching unit surface area of the edges and surfaces of the plates will be the same, so that as the plate grows the ratio of its thickness to its breadth should increase. In his treatment Nabarro finds that the dynamical equilibrium ratio of the thickness to the breadth of a large grain is proportional to the supersaturation, so the higher the degree of supersaturation, the greater will be the strain.

These considerations may be utilized to suggest an explanation of the fact that in certain circumstances precipitation is discontinuous and in others continuous.

Discontinuous precipitation. The experimental results concerning discontinuous precipitation are summarized below:

(1) A crystal in which discontinuous precipitation is taking place is composed largely of regions in which precipitation either has not commenced or is complete, with few regions in which precipitation is in an intermediate stage.

(2) Discontinuous precipitation occurs when the degree of supersaturation is high and the temperature is low.

(3) In all silver-rich and copper-rich single crystals examined, if precipitation is discontinuous, fragmentation occurred simultaneously with precipitation.

(4) From a survey of the literature on the subject, Dehlinger (1939) has pointed out that discontinuous precipitation occurs only in alloy systems where there is a large difference between the atomic sizes of the solvent and precipitate.

Thus it is seen in (2) and (4) that discontinuous precipitation occurs when, according to the theoretical work of Nabarro, large strains are likely to be produced during the precipitation process, and it seems possible that the elastic energy associated with the growing nuclei may reach such a high value that the matrix in the neighbourhood of the precipitate collapses under the strain. In the resulting conglomeration of distorted crystallites there should be many places suitable for the formation of further nuclei of precipitate, so that precipitation may become catastrophic in the region. Once precipitation commences in a region, therefore, it will proceed very rapidly to completion in that region.

The interpretation is in accordance with all the experimental results tabulated above. It follows, as a consequence of this suggested mechanism for discontinuous precipitation, that in all alloy systems where precipitation of this type occurs, it should be accompanied by collapse of the matrix and hence a fragmentation of the solid solution.

Continuous precipitation. Continuous precipitation occurs when the degree of supersaturation is low, and therefore according to Nabarro the strain associated with the growing nuclei is likely to be small. Experimentally it is found that fragmentation either does not occur in discontinuous precipitation or if it does take place, then it commences only towards the end of the precipitation process. This result is thus in agreement with the hypothesis that fragmentation is associated with high strain.

The strains produced by precipitation under low degrees of supersaturaretion a kept small since the precipitate can take up a plate-like form which has a low elastic energy (Nabarro 1940).

There are, however, some difficulties confronting the viewpoint that plate formation is the sole mechanism by which strain is relieved when continuous precipitation occurs with no break-up of the single crystals.

The copper-rich solid solution contains about 1·4 times as many atoms per unit volume as the silver-rich solid solution, so that if the elastic energy is to be kept low, atoms must be moved during precipitation by some other means than merely place-changing. The orientation of the precipitate is the same as that of the matrix. It might be expected that when the precipitate breaks away from the matrix and shifting of the atoms occurred, the orientation was likely to be upset. Furthermore, unless the plate of precipitate extends right across the crystal (or possibly across the mosaic crystallite in the case of a single crystal) some fragmentation is likely in order to relieve the elastic strains which will be high in the neighbourhood of the edges of the plate. In an alloy containing 4·0 % silver no fragmentation was observed, although it was held at the ageing temperature (500° C) for a long period after precipitation was complete.

It seems probable, therefore, that diffusion of vacant lattice sites or interstitial atoms may also play a part in the removal of elastic strains. The number of lattice

defects (vacant sites or interstitial atoms) in equilibrium increases with temperature, so that it is to be expected that the relief of elastic strain by this means would become more effective as the temperature increases, and precipitation should tend to become continuous. Although the results given in table 2 appear to agree with this deduction, it should be noted that comparison has not been made at two different temperatures of two alloys having the same degree of supersaturation at their respective ageing temperatures. It is therefore possible that the differences in behaviour may be due to the different supersaturations.

The experimental results do not permit an estimate to be made of the relative importance of plate formation and diffusion of lattice defects as mechanisms for reducing elastic strains.

REFERENCES

Ageew, Hanson & Sachs 1930 *Z. Phys.* **66**, 350.
Ageew & Sachs 1930 *Z. Phys.* **63**, 293.
Barrett & Geisler 1940 *J. Appl. Phys.* **11**, 733.
Barrett, Kaiser & Mehl 1935 *Ann. Instn Mech. Engrs*, **117**, 39.
Becker 1937 *Z. Metallk.* **29**, 245.
Becker 1938 *Ann. Phys., Lpz.*, **32**, 128.
Becker 1940 *Proc. Phys. Soc.* **52**, 71.
Bumm & Dehlinger 1936 *Metallwirtsch.* **15**, 89.
Cohen 1937 *Ann. Instn Mech. Engrs*, **124**, 138.
Cox & Sykes 1940 *J. Inst. Met.* **66**, 381.
Dehlinger 1939 *Chem. Phys. der Metalle und Legierungen.* Leipzig.
Jones & Leech 1941 *J. Inst. Met.* January.
Mehl, Barrett & Rhines 1932 *Ann. Instn Mech. Engrs*, **99**, 203.
Mehl & Jetter 1939 *Amer. Soc. Met.* Preprint no. 30.
Mott & Nabarro 1940 *Proc. Phys. Soc.* **52**, 87.
Nabarro 1940a *Proc. Phys. Soc.* **52**, 90.
Nabarro 1940b *Proc. Roy. Soc.* A, **175**, 519.
Owen & Rogers 1935 *J. Inst. Met.* **57**, 257.
Preston 1938 *Proc. Roy. Soc.* A, **167**, 526.
Rosenhain 1908 *J. Iron Steel Inst.* **1**, 87.
Schmid & Siebel 1933 *Z. Phys.* **85**, 36.
Schmid & Siebel 1934 *Metallwirtsch.* **13**, 765.
Wiest 1932 *Z. Phys.* **74**, 225.
Wiest 1933 *Z. Metallk.* **25**, 238.
Wiest & Dehlinger 1934 *Z. Metallk.* **26**, 150.

On nuclear energy levels

By K. M. Guggenheimer

The formula for the energy levels of the rigid rotator has been applied to nuclei. Two kinds of nuclear rotation are discussed: the rotation of the whole nucleus leads to an $A^{-\frac{1}{3}}$ relation for the energy levels of different nuclei; the rotation around the nucleus leads to an $A^{-\frac{2}{3}}$ relation. Both relations are based on the assumption of equal density of different nuclei. Examples for both relations have been found.

The existence of simple rational relations between the excitation levels of a nucleus, predicted by the rotator formula and empirically known in some heavy nuclei, has been confirmed in other nuclei.

The value of the rotation constant has been determined: $B_0 = 5 \cdot 06_4$ MeV.

The system of rotator levels appears to be relevant to the interpretation of the observed nuclear levels.

The transition probabilities in a system of rotator levels is discussed. Nuclear excitation levels are compared with X-ray levels.

From the level system can be concluded that the nuclear radii fulfil the relation $r = r_0 A^{\frac{1}{3}}$ with $r_0 = 2 \cdot 03_1 \, 10^{-13}$ cm. This value is in conformity with values found by other methods.

Twenty years ago, Ellis (1922) suggested, for the first time, a system of discrete nuclear energy levels for describing the origin of a system of γ-rays. He discussed the applicability of the combination rule for these rays and wrote: 'The general results lend support to the view that stationary states exist in the nucleus and it is hoped that by investigating the gamma-rays of other radioactive bodies some general principles in the arrangement of nuclear levels may be found.' The assumption of the existence of discrete nuclear excitation levels was confirmed later on by the observation of the fine structure of α-rays and of corresponding well-defined γ-rays. Measurements of the energy release in artificial nuclear reactions indicated the occasional excitation of nuclei in these reactions and provided more data for nuclear levels. But the hope of finding some general principles for the values of nuclear levels has not been fulfilled up to now.

The quantum mechanical prediction of nuclear energy levels is not possible as long as the law of interaction between the particles which constitute the nucleus is not known. On the assumption of an exponential law, levels of light nuclei have been calculated by Feenberg and Wigner and others, and general considerations on the symmetry of the Hamiltonian allowed some inferences as to the possible grouping and spacing of energy levels (cf. Wigner & Feenberg 1941). It is not possible to say that these predictions have so far been fulfilled (cf. Margenau 1941).

Bohr & Kalckar (1937) showed that vibrations of the whole nucleus should vary as $A^{-\frac{1}{2}}$, where A is the atomic weight, and that at least the lowest known levels do not correspond to vibrations of this kind. Vibrations of the surface should lead to higher energy levels than vibrations of the whole nucleus; they should vary as $A^{-\frac{1}{2}}$, and are expected to possess energies of more than 1 MeV even for

the heaviest nuclei. But empirically the variation is apparently more rapid than $A^{-\frac{1}{3}}$ and the energy of the lowest nuclear levels which are known is of the order 10^4 eV.

Apart from vibrations, there is the possibility that rotational motions may cause nuclear energy levels. Terms due to the rotation of the whole nucleus should vary as $A^{-\frac{5}{3}}$. Teller & Wheeler (1938) discussed theoretically the level system due to nuclear rotation, but came to the conclusion that no low-lying levels of this kind could be expected in nuclei. Still, the order of values, computed by them for lead, would correspond quite well with the order of terms known in radioactive nuclei, e.g. ^{223}AcX.

A principal reason, however, for taking up again the discussion of the rotational formula is the following: It is known that the nuclear forces of attraction arise quite suddenly near the nucleus, and the curve describing the potential must be rather steep. Now, the virial theorem of Clausius, which holds also in quantum mechanics, gives, for stationary states, a relation between the mean kinetic energy T and the mean potential energy $V = Cr^n$, depending upon the power of the mean distance r of the particles

$$T = \tfrac{1}{2} n V. \tag{1}$$

The total energy E of the system can be expressed in terms of the mean kinetic energy of the system

$$E = T + V = (1 + 2/n)\, T. \tag{2}$$

The higher the power n, the more will the total energy consist mainly of the kinetic energy. In particular, if the system becomes excited by increasing the total energy by a finite amount, the potential energy will only change by the fraction $2/n$ of T. In the limit $n = \infty$, the energy of excitation consists only of kinetic energy, and since then the mean distance does not vary at all, the condition for a rigid rotator is fulfilled. The steeper the potential curve at a given distance, the better the approximation by the formula of the rigid rotator.

On the other hand, even in the case of molecules, with forces of attraction much smaller than in nuclei, the formula of the rigid rotator is a very good approximation for representing the closely spaced molecular energy levels. Therefore in nuclei with their strong forces of attraction, rotational motions may very well occur. It remains a question of empirical research to find out whether examples can be found where nuclear levels obey the rotator formula.

THE TWO KINDS OF ROTATION

In non-relativistic quantum mechanics the eigenvalues E of the rigid rotator are given by

$$E = B \times K(K+1) \tag{3}$$

with

$$B = \hbar^2/2I = \hbar^2/2Mr^2, \tag{4}$$

where $2\pi\hbar$ is Planck's constant, I is the moment of inertia, r is the radius of the rotator and M is the rotating mass. If the whole nucleus rotates and has spherical symmetry, the moment of inertia is

$$I = 0 \cdot 4 M r^2 = 0 \cdot 4 M_0 A r^2, \tag{5}$$

where A is the mass number and M_0 is the mass which corresponds to unit atomic weight.

From considerations of binding energies and of cross-sections it seems to be a good approximation to suppose that the volume of nuclei increases proportionally to the mass number A, i.e. to the number of particles in the nucleus. Then

$$r = A^{\frac{1}{3}} r_0. \tag{6}$$

Under this condition, (4) assumes the form

$$B_A = 2 \cdot 5 B_0 A^{-\frac{1}{3}} \tag{7}$$

with

$$B_0 = \hbar^2 / 2 M_0 r_0^2, \tag{8}$$

and the eigenvalues of the energy are obtained by inserting (7) and (8) into (3)

$$E = 2 \cdot 5 B_0 A^{-\frac{1}{3}} K(K+1). \tag{9}$$

Apart from the rotation of the whole nucleus, there is another possible origin of rotational energy levels, namely, the rotation of a single, or a few, particles around the nucleus. The closest approach of an additional particle is equal to the radius of the compound nucleus. Also during the rotation, the rotating particles have to remain within the range of attraction, which falls off very quickly with the distance. The radius of rotation, therefore, may be considered to be the same as the radius of the compound nucleus. In this case, the rotation of a single particle around the nucleus obeys the formula (3) with

$$B = B_0 A^{-\frac{1}{3}}. \tag{10}$$

The rotation of several particles leads to

$$B = B_0 A^{-\frac{1}{3}} a^{-1}, \tag{11}$$

and thus to the energy levels

$$E = B_0 A^{-\frac{1}{3}} K(K+1) a^{-1}. \tag{12}$$

where a is the number of particles rotating around the nucleus.

Given the radius, the eigenvalues of the energy would be determined by (6), (8) and (9) or (12). Conversely, the empirical determination of B would provide knowledge of the radius. The question of the size of nuclear radii will be discussed in the last section. Various methods for the determination of the nuclear radii do not yet appear to have arrived at an exact and unambiguous result. Therefore, for the present, it has to be investigated, whether the value of B_0 or of any B may be determined empirically by the analysis of nuclear levels or nuclear spectra.

The determination of B_0

The best cases to be analysed are those where many levels are known and carefully measured. This condition is fulfilled for some heavy radioactive bodies, e.g. ^{223}AcX. The magnetic analysis of Rosenblum (1937) of the fine structure of α-rays of ^{227}RaAc indicated levels of ^{223}AcX which can be arranged into two series with similar values of B (cf. table 1) (Guggenheimer 1940).

Table 1. AcX

$E_{obs.}$	$E_{calc.}$		$E_{obs.}$	$E_{calc.}$	
386	382_7 kV	$K\,15$	197	197_1 kV	$K\,11$
335	335_0	$K\,14$	174	175_4	$K\,10$
313	313_6	$K\,14$	135	134_4	$K\,9$
304	——	—— line of An?	(120)	114_8	$K\,8$ weak line
288	290_1	$K\,13$	(96)	89_3	$K\,7$ weak line
(260)	271_8	$K\,13$ weak line	81	83_6	$K\,7$
247	248_7	$K\,12$	61	62_7	$K\,6$
235	233_0	$K\,12$	30	29_9	$K\,4$
(217)	210_5	$K\,11$ weak line			

The values of B are calculated by representing the empirical values of the energy levels by the rotator formula with least square deviations. If only the intense rays are used, one obtains for the two series $B = 1.49_{35}$ and $B = 1.59_{45}$ kV and as mean value $B = 1.54_4$ kV.

If one inserts this value into (7), one obtains

$$B_0 = 5.06_4 \text{ MeV.} \qquad (13)$$

The second example was found in ^{26}Mg with levels at 2·3, 4·0 and 5·0 MeV (May & Vaidyanathan 1936). The last two levels show at once the simple rational relation which is to be expected for rotator terms. The ratio $4:5$ corresponds to levels with $K = 8$ and $K = 9$. This leads to $B = 55.5$ kV and with (7) to $B_0 = 5.06_6$ MeV, which is very similar to the value $B_0 = 5.06_4$ MeV found with ^{223}AcX.

The third value 2·3 MeV in ^{26}Mg compares with 2·33 MeV with $K = 6$ in (9). Thus all three values correspond to levels which are to be expected by the analysis of the levels of ^{223}AcX.

A second possibility of determining B_0 or of testing (13) is provided by the $A^{-\frac{1}{3}}$ relation (11). This relation is to be applied particularly to high levels and to large level spacing.

Rutherford & Lewis (1933) have already emphasized the existence of simple numerical relations within the level system of ^{214}RaC′. They give for twelve excitation levels the formula $1414 + p426 + q182$ kV. The last value corresponds exactly to a level with $K = 10$ with (9) and (13). The two first values which represent also the most intense α-ray and a γ-ray, show the ratio $10:3$.

If one takes as the largest common factor F, the unit for which the deviations of corresponding multiples show least square deviations from the experimental

values, one obtains $F = 141 \cdot 4_5$ kV. This value should be in a simple rational relation to $B = B_0 A^{-\frac{1}{3}}$ according to (10). Inserting (13) and $A = 214$, one obtains $B = 141 \cdot 5_3$.

Another nucleus whose levels have been carefully measured is ^{208}ThC'' (Rosenblum 1937). Some level distances revealed by α-rays and γ-rays show, apart from small level spacings, definite multiples of 144 kV, namely, $\gamma 287 = \alpha\gamma 327 - \alpha\gamma 40 \simeq 2 \times 144$ kV; $\gamma 432 = \alpha\gamma 472 - \alpha\gamma 40 = 3 \times 144$ kV; $577 = \gamma 617 - \alpha\gamma 40 \simeq 4 \times 144$ kV. According to (10) and (13), $B = B_0 208^{-\frac{1}{3}} = 144 \cdot 2$ kV.

Burcham & Smith (1938) found with $(d\alpha)$ reactions, in ^{17}O four levels at $0 \cdot 83$, $2 \cdot 95$, $3 \cdot 77$ and $4 \cdot 49$ MeV, nearly within the ratio $1 : 4 : 5 : 6$. The unit F is ~ 750 kV. With (10) and (13) and $A = 17$, one finds $B = B_0 17^{-\frac{1}{3}} \simeq 766$ kV, and corresponding levels at $0 \cdot 77$, $3 \cdot 07$, $3 \cdot 83$ and $4 \cdot 60$ MeV.

According to the scheme of rotator levels, the energies of γ-rays of a nucleus should show rational proportions, if they arise from this system. The level scheme derived from γ-rays alone cannot provide the same certainty as the levels found in nuclear reactions, but the differences of the energies should lead to simple numerical relations with the value of B in (9) or (11), if the nucleus is spherically symmetrical.

Among recent measurements may be mentioned four hard γ-rays with $6 \cdot 1$, $6 \cdot 5$, $7 \cdot 7$ and $8 \cdot 2$ MeV in ^{28}Si (Plain, Herb, Hudson & Warren 1940). The values are nearly in the ratio $11 : 12 : 14 : 15$. The largest common factor $F = 548$ kV compares with $B = B_0 28^{-\frac{1}{3}} = 549$ kV.

In ^{82}Br three γ-rays have been reported (Roberts, Downing & Deutsch 1941) with $0 \cdot 547$, $0 \cdot 787$ and $1 \cdot 35$ MeV. They show practically the ratio $2 : 3 : 5$. The largest common factor $F = 268 \cdot 4$, and $B = B_0 82^{-\frac{1}{3}} = 268 \cdot 3$ kV.

All these examples can be considered as evidence in favour (1) of the occurrence of rational relations between energy levels of a nucleus, $(2a)$ of the applicability of the $A^{-\frac{1}{3}}$ relation (7), or $(2b)$ of the applicability of the $A^{-\frac{1}{3}}$ relation (10) which both follow from the assumption of equal nuclear density, (3) of the value (13) of B_0 (which has been found originally by the $A^{-\frac{1}{3}}$ relation alone).

On this basis, it is possible to attempt the interpretation of the levels themselves and to attribute to them certain quantum numbers.

Interpretation of nuclear levels

With the value (13) $B_0 = 5 \cdot 06_4$ MeV, the level scheme of figure 1 has been constructed.

The abscissa gives the atomic weight between 8 and 238 on a logarithmic scale, the ordinate shows the energy between 20 kV and 20 MeV, also on a logarithmic scale.

The $A^{-\frac{1}{3}}$ relation (9) and the $A^{-\frac{1}{3}}$ relation (12) give the two systems of parallel straight lines and show the theoretical position of the levels.

The black dots signify the nuclear levels which have been determined empirically.

The three highest levels of ^{17}O lie on three consecutive lines, and the four levels measured by Bower & Burcham (1939) in (dp) ^{20}F with 0·7, 1·0, 1·35 and 1·9 MeV lie on, or near, four consecutive levels (1, 1½, 2, 3) of the A^{-1} relation (12).

There are several nuclei where all measured levels seem to belong to the same a series, even to consecutive levels of the series, e.g. the three levels of ^{34}S, the two levels of ^{42}Ca and the two of ^{48}Sc ($a2$). The two levels of ^{41}A lie on two consecutive lines of the $a4$ series.
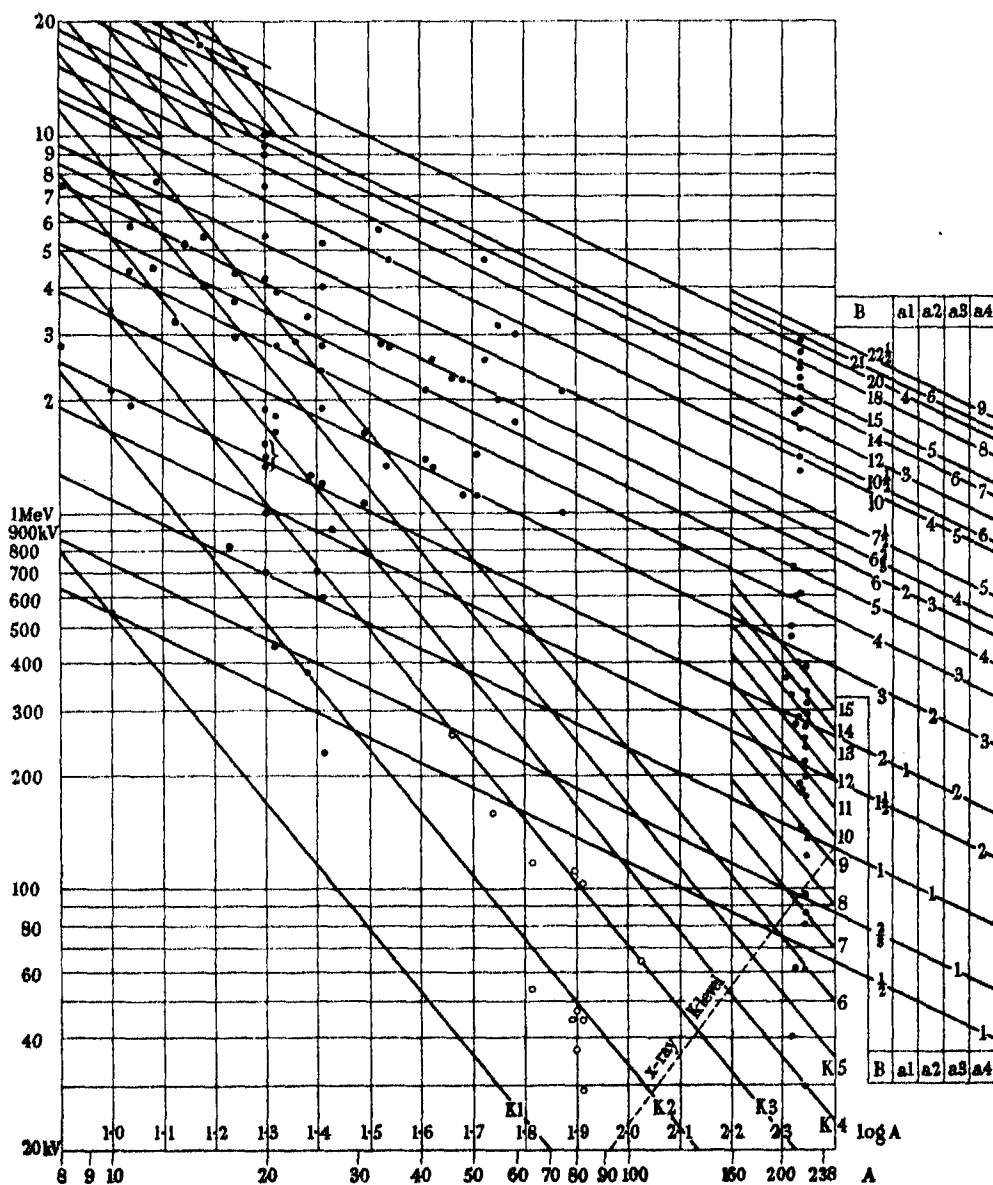
FIGURE 1. ● energy levels; ○ γ-rays.

The cases where all observed levels of a nucleus lie on consecutive lines of the $A^{-\frac{1}{3}}$ relation, or belong to the same a series, may be considered as evidence in favour of the theoretical level system.

The crossing of two lines may be connected with an increased probability of the excitation of the corresponding energy. The following crossing points are marked by dots which represent observations: $K1$ with $a4K1$; $K3$ with $a4K2$; $K4$ with $a1K1$; $K5$ with $a4K4$; $K6$ with $a3K3$ and with $a2K3 = a1K2$.

The highest nuclear excitation observed up to the present is 17·5 MeV (cf. figure 1). It lies on the line $21B$ and is therefore interpreted as $a2K6$ level. Empirically, it is a deuteron-resonance level in ^{15}N (Bennett, Bonner, Hudspeth & Watt 1941), and the interpretation here is in agreement with this fact.

The highest level observed in the α-ray fine structure of heavy nuclei is 2·880 MeV in RaC'. It lies above the line $20B$. The highest γ-ray of the heavy radioactive nuclei has been found to be 3·2 MeV in ThD, which would lie on the line $22\frac{1}{2}B$, $a4K9$.

The position of the lowest nuclear levels is of particular interest. Theoretically, the lowest level is at 2·770 kV in ^{238}U. The lowest observed level in α-radioactive nuclei is 31 kV in ^{223}AcX. It corresponds to a $K4$ level. The lowest γ-rays reported in the list of Seaborg (1940) are 54 kV in $A = 65$, 46 kV in $A = 78$; 37 and 49 kV in $A = 80$ and 29 and 46 kV in $A = 83$. Three of these lines correspond to the energy of the $K2$ level (53 kV in $A = 78$, 51 in $A = 80$ and 48 in $A = 83$); these lines can also be due to transitions to $K3$–$K2$. But the other three lines can, within the level scheme, only be attributed to the transitions $a2$–$K1$. (Within this scheme, they cannot represent the level $K1$, because then, the next higher rays would possess twice as much energy.)

The smallest energy released by nuclei, has been reported in ^{134}Ba (Collar, Cork & Smith 1941), namely, a γ-ray with 17 kV, whereas 14·4 kV corresponds there to $K2$–$K1$.

It is noteworthy that nuclear rotation leads to such low-lying levels at these atomic weights.

In ^{7}Li, a level is known to lie at 0·5 MeV (Wilson 1941). If this level is due to rotation of the whole nucleus, the nucleus must be rather flat (perhaps a centred hexagon).

The level system of ^{208}ThC'' is of special interest. The transition to the first excitation level of 40 kV is more frequent than to the ground level. The value of 40 kV is too small to be part of the $A^{-\frac{1}{3}}$ system (12). But there are combinations of energies of this system ($B = 144\cdot2$ kV) with the first value, which are excited and irradiated simultaneously. Here the interpretation is that the excitation of the rotation of the whole nucleus can simultaneously occur with a rotation around the nucleus, and more often the de-excitation occurs separately.

Multiple excitation is also indicated by the formula of Rutherford & Lewis for the levels of ^{214}RaC'.

Resonance levels correspond generally to such high energies of excitation that

the probability of multiple excitation is rather great. Thus, these levels are less favourable for the determination of the fundamental level system.

The occurrence of multiple excitation forms an essential part in Bohr's theory of the compound nucleus (Bohr 1936). In fact, the level density found in resonance levels is much higher than the level density observed with nuclear reactions, viz. the differences of the released energies. Thibaud & Comparat (1939) found, in ^{15}N, fourteen resonance levels between 11·64 and 14·16 MeV, Stuhlinger (1939) in the same nucleus nineteen resonance levels between 14·64 and 17·45 MeV. Fünfer (1939) reported twenty-two resonance levels in ^{13}C between 12·67 and 16·68 MeV; also in ^{18}F, ^{31}P and ^{44}Ca, eleven, thirteen and sixteen resonance levels have been found.

Bohr's idea of multiple nuclear excitation can be applied also to the system of rotator levels. As mentioned in the case of ^{208}Th C″, the rotation of the whole nucleus can be excited simultaneously with the rotation of a group of particles around the nucleus. If one restricts the permission of multiple excitation and if only a double excitation of this kind is allowed, namely, the combination of the excitation of a level of the $A^{-\frac{4}{3}}$ series together with one of the four a series of the $A^{-\frac{1}{3}}$ relation, then one obtains already at least the same level density as the observed one (even if the rotator levels remain $2K + 1$-fold degenerate).

MULTIPOLE RADIATION

The selection rule, which is valid for transitions between rotator levels of molecules, cannot be expected to hold in nuclei. The observed values of the rotation constant B of diatomic molecules lie between

$$4·631 \times 10^{-6} \text{ eV (I}_2) \quad \text{and} \quad 7·545 \times 10^{-3} \text{ eV (H}_2),$$

whereas for nuclei the value $B_0 = 5·064 \times 10^6$ eV has been adopted.

The emission probability of photons depends upon the energy and the multipole order, according to the formula (Segré; cf. Helmholtz 1941)

$$\lambda = \left(\frac{\omega}{c}\right)^{2l} \frac{c^2}{\hbar} \eta \frac{x^{2l}}{[1! \, 3! \, \dots \, (2l-1)!]^2}, \tag{14}$$

where λ is the decay constant, $\hbar\omega$ the energy, η a constant of the order 1 and x is of the order of the nuclear radius. The emission probability increases with the energy of the photons, and decreases with the multiple order of the emitted light. But in a system of rotator levels, the multipole order increases with the energy according to

$$E = B[K(K+1) - (K-l)(K-l+1)] = B[(2K+1)l - l^2]. \tag{15}$$

It depends, therefore, apart from the nuclear charge, upon the values of B and of K, whether the transition to the ground level or to the neighbouring level or to an intermediate level will be the most probable.

The occurrence of γ-radiation of high multipole order (up to 2^5) has been observed (cf. Helmholz 1941).

NUCLEAR LEVELS AND X-RAY LEVELS

In figure 1, a line has been drawn at the lower right-hand corner, which connects the points $A = 92$, $E = 20$ kV, and $A = 238$, $E = 115.3$ kV. It indicates the region of the X-ray K-level.

According to (9), from ^{86}Rb on, nuclear levels should lie lower than the X-ray K-level of the same atom. The occurrence of nuclear energy levels below the X-ray K-level is established in heavy radioactive bodies (cf. figure 1).

It is easily seen in figure 1, that, for example, the energy of 115.3 kV of the uranium X-ray K-level should be sufficient for the excitation of the nuclear level $K1$ of $A = 26$. This nucleus should show a selective absorption of radiation of this energy, and the absorption of this energy should be larger in this nucleus than in nuclei of smaller or higher atomic weight. Again, a nucleus with the atomic weight $A = 50$, should show a similar selective absorption owing to the nuclear level $K2$. The excitation of a nuclear level with $K > 1$ should give rise to the emission of secondary rays of greater wave-length.

Systematic absorption measurements of this kind would provide a simple method for testing the position, in particular, of the lowest nuclear energy levels.

NUCLEAR RADII

The $A^{-\frac{1}{3}}$ relation (12) as well as the $A^{-\frac{1}{3}}$ relation (9) are based on the assumption of equal nuclear density, $r = A^{\frac{1}{3}} r_0$. Equation (8), $B_0 = \hbar^2/2 M_0 r_0^2$, together with the empirical value (13) of $B_0 = 5.064$ MeV, allows the evaluation of r_0. With $2\pi \hbar = 6.6242 \times 10^{-27}$ erg sec. and $M_0 = 1.660 \times 10^{-24}$ g., one obtains

$$r_0 = 2.031 \times 10^{-13} \text{ cm.} \tag{16}$$

This value, derived from the analysis of nuclear levels, is in remarkable agreement with the value 2.05×10^{-13} cm. deduced by Bethe (1937) by the theory of α-decay under the assumption that α-particles are not pre-existent in the nucleus. With this assumption he was led to correct the value $r_0 = 1.4 \times 10^{-13}$ cm., assumed by Gamow (1931/7).

Recently, proton-proton, neutron-proton and neutron-deuteron scattering experiments have been described with a range of forces between 1.73×10^{-13} and 2.8×10^{-13} cm. (e.g. Breit, Thaxton & Eisenbud 1939; Rarita & Schwinger 1941; Buckingham & Massey 1941).

Kuerti & Wilkins (1940) measured the scattering of protons in aluminium and found as the limit of the validity of the Coulomb law a radius of 6.8×10^{-13} cm., whereas (6) and (16) lead to $r = 6.09 \times 10^{-13}$ cm.

Placzek & Bethe (1940) remarked that the formula for Fraunhofer scattering can be applied to nuclei. By this method Present (1941) computed from recent scattering experiments of Wakatuki (1940) with neutrons, radii for iron $r_{\text{Fe}} = 7.4 \times 10^{-13}$ cm. and for lead $r_{\text{Pb}} = 11.3 \times 10^{-13}$ cm. whereas with (16) one has

to expect rotator radii of iron $r_{Fe} = .7 \cdot 6 \times 10^{-13}$ cm. and of lead $r_{Pb} = 11 \cdot 7 \times 10^{-13}$ cm. The scattering values of small angles seem to lead to even higher values of the radii.

Barschall & Ladenburg (1942) measured the total cross-sections for the elastic and inelastic scattering of $2 \cdot 5$ MeV neutrons in several nuclei (cf. table 2). The de Broglie wave-length of these neutrons is $1 \cdot 82 \times 10^{-12}$ cm. Even for light nuclei, the empirical values of the cross-sections q are larger than $\lambda^2/\pi = 1 \cdot 05 \times 10^{-24}$ cm.[2] (cf. Peierls 1940).

One can define an effective scattering radius r_s by the formula

$$q = r_s^2 \pi. \tag{17}$$

The differences between the scattering radii, derived from the observations of Barschall & Ladenburg, and the geometrical radii (16) are on the average $2 \cdot 0 \times 10^{-13}$ cm. Thus, the scattering radii can approximately be written in the form

$$r_s = r + r_0 = r_0(A^{\frac{1}{3}} + 1), \tag{18}$$

and are practically equal to the sum of the radii of the nucleus and the neutron. The cross-sections q_c calculated with (18), (17), (16) are inserted in table 2.

TABLE 2. CROSS-SECTIONS

	C	Al	Fe	Cu	Zn	Pb
$q_{obs.} \times 10^{-24}$	$1 \cdot 6 \pm 0 \cdot 3$	$2 \cdot 4 \pm 0 \cdot 3$	$3 \cdot 1 \pm 0 \cdot 3$	$2 \cdot 7 \pm 0 \cdot 3$	$3 \cdot 0 \pm 0 \cdot 3$	$6 \cdot 0 \pm 0 \cdot 7$
$q_{calc.} \times 10^{-24}$ cm.[2]	$1 \cdot 40$	$2 \cdot 07$	$3 \cdot 02$	$3 \cdot 24$	$3 \cdot 27$	$6 \cdot 18$

(With a value $r_0 = 1 \cdot 47 \times 10^{-13}$ cm., the differences between the scattering radii and the geometrical radii would be on the average greater than 4×10^{-13} cm.)

Pollard, Schultz & Brubacker (1938) found for the reaction radii of neon, aluminium and argon values which correspond at least to $r_0 = 1 \cdot 94 \times 10^{-13}$ cm. Weisskopf & Ewing (1940) computed a value $r_0 = 1 \cdot 3 \times 10^{-13}$ cm. for the reaction radii. It is the value of the sticking probability which makes this method rather uncertain (cf. Dunlap & Little 1941).*

Weizsäcker (1935) deduced with his semi-empirical formula for packing fractions and isobaric shift, a value $r_0 \cong 1 \cdot 4 \times 10^{-13}$ cm. Bethe showed (1937) that the semi-empirical formula can be adapted to the larger radius $r_0 = 2 \cdot 05 \times 10^{-13}$ cm.

---

* Krishnan & Nahum (1942) measured cross-sections and excitation functions of (dp) and (dn) reactions of heavy nuclei. Owing to the uncertainty of the sticking probability, the absolute values of the cross-sections seem to be compatible with $r_0 = 2 \cdot 0 \times 10^{-13}$ cm. as well as with $r_0 = 1 \cdot 47 \times 10^{-13}$ cm. The cross-sections of the (dp) reactions have been found to be much larger than those of the (dn) reactions. The Oppenheimer-Phillips process provides a fitting factor, if $r_0 = 1 \cdot 47 \times 10^{-13}$ cm. or even somewhat higher. On the other hand, the observed facts can be interpreted and are compatible with $r_0 = 2 \cdot 0 \times 10^{-13}$ cm., and $\Gamma_p > \Gamma_n$, where Γ_p and Γ_n are the emission probabilities of a proton or of a neutron. It also should be kept in mind that calculations of penetrabilities become increasingly less reliable as the bombarding energy approaches values which are comparable with the height of the barrier (cf. Volkoff 1940).

Aten (1939) found that with very simple assumptions for the surface energy and binding energies, the isotopic shift leads to a value $r_0 = 2 \cdot 05 \times 10^{-13}$ cm. (cf. also Jordan 1937).

Nuclear radii have been calculated by using the maximum energies of β-rays. The nuclei with $Z = N + 1$, in particular, show a remarkable regularity and lead to an $A^{-\frac{1}{3}}$ relation of the form (6). Under the assumption of uniform distribution of the nuclear charge, a value $r_0 = 1 \cdot 47 \times 10^{-13}$ cm. has been derived for the 'Coulomb-radii' (Wigner 1937; Bethe 1938; Stephens 1940; Elliot & King 1941).

The charge emission causes in the remaining nucleus the change of the Coulomb field, and it is generally supposed that this change is the only alteration of the energy of interaction (at least in the series $Z = N + 1$). On the other hand, the energy release allows the calculation of the change of the total energy in the nucleus. If this change of the total energy consists only of, and is equal to, the change of the potential energy, the radii can be calculated with the result mentioned above.

But equation (1) shows, as a result of the virial theorem, that the change of the total energy is generally not at all equal to the change of the potential energy. The same is true in quantum mechanics. The equality of the change of the eigenvalue of the energy with the change of the potential energy, treated as a perturbation, would be only a first-order approximation. The second-order approximation cannot be carried out as long as the exact law of interaction in nuclei and the eigenfunctions are not known. The inference from the change of the total energy to the radius is therefore rather hypothetical. If, for example, the rotator radius $r_0 = 2 \cdot 03 \times 10^{-13}$ cm. is correct, it can be concluded that the ratio between the change of the total energy and of the potential energy is $\Delta E \simeq 1 \cdot 4 \Delta V$ for the ground state. Thus, the large radius is not incompatible with the observed values of the energy releases in β-radiation.

Nordheim & Yost (1937) mentioned already that the large radius does not affect essentially the consequences of Fermi's theory of β-decay.

In table 3 the values of r_0 calculated with different methods, are assembled:

TABLE 3. NUCLEAR RADII

α-decay radius	Gamow 1931/7	Bethe 1937
	$r_0 = 1 \cdot 4 \times 10^{-13}$ cm.	$r_0 = 2 \cdot 05 \times 10^{-13}$ cm.
Scattering radius		Barschall & Ladenburg 1942
		$r_0 > 2 \times 10^{-13}$ cm.
Reaction radius	Weisskopf & Ewing 1940	Pollard et al. 1938
	$r_0 = 1 \cdot 3 \times 10^{-13}$ cm.	$r_0 = 1 \cdot 94 \times 10^{-13}$ cm.
Semi-empirical radius	Weizsäcker 1935	Jordan 1937
	$r_0 \simeq 1 \cdot 4 \times 10^{-13}$ cm.	$r_0 = 2 \cdot 14 \times 10^{-13}$ cm.
		Aten 1939
		$r_0 = 2 \cdot 05 \times 10^{-13}$ cm.
Coulomb radius	Wigner 1937, Bethe 1938	
	$r_0 = 1 \cdot 47 \times 10^{-13}$ cm.	$r_0 = 2 \cdot 03 \times 10^{-13}$ cm.
	if $\Delta E = \Delta V$	if $\Delta E = 1 \cdot 4 \Delta V$
Rotator radius		$r_0 = 2 \cdot 03_1 \times 10^{-13}$ cm.

The rotator radius agrees well with the α-decay radius, with the recent determinations of the scattering radius and of the semi-empirical radius of binding energies and isobaric shift of stability. It is compatible with the reaction radius and with the Coulomb radius.

Improvements of the determination of nuclear rotation levels would allow one to determine the nuclear radii with the same spectroscopic exactitude as the radii of diatomic molecules.

The value of r_0 is not identical with the radius of a single particle. If the particles of a nucleus are distributed like most densely packed spheres, the radius of each of these spheres is $0 \cdot 90 r_0$. With other distributions, the radius of the particles comes out even smaller than this value.

The value (16) of r_0 can be considered as evidence in favour of the interpretation of the level system (9) and (12) as due to nuclear rotation (cf. table 3).

REFERENCES

Aton, A. H. W. 1939 *Physica*, 6, 425.
Barschall, H. H. & Ladenburg, R. 1942 *Phys. Rev.* 61, 129.
Bennett, W. E., Bonner, T. E., Hudspeth, H. T. & Watt, B. E. 1941 *Phys. Rev.* 59, 781.
Bethe, H. A. 1937 *Rev. Mod. Phys.* 9, 69.
Bethe, H. A. 1938 *Phys. Rev.* 54, 436.
Bethe, H. A. 1940 *Phys. Rev.* 57, 260, 390, 1125.
Bohr, N. 1936 *Nature, Lond.*, 137, 344.
Bohr, N. & Kalckar, F. 1937 *K. danske vidensk. Selskab. Math.-fys. Medd.* 14, Nr. 10.
Bonner, T. W. 1940 *Proc. Roy. Soc.* A, 174, 339.
Bonner, T. W. & Brubaker, W. M. 1936 *Phys. Rev.* 50, 308.
Bower, J. C. & Burcham, W. E. 1939 *Proc. Roy. Soc.* A, 173, 379.
Breit, G., Thaxton, M. & Eisenbud, L. 1939 *Phys. Rev.* 55, 1018.
Buckingham, R. A. & Massey, H. S. W. 1941 *Proc. Roy. Soc.* A, 179, 123.
Burcham, W. E. & Smith, C. L. 1938 *Proc. Roy. Soc.* A, 168, 176.
Collar, H. W., Cork, J. M. & Smith, G. P. 1941 *Phys. Rev.* 59, 937.
Davidson, W. 1939 *Phys. Rev.* 56, 1062.
Davidson, W. 1940 *Phys. Rev.* 57, 568.
Davidson, W. & Pollard, E. 1938 *Phys. Rev.* 54, 408.
Dunlap, H. F. & Little, R. N. 1941 *Phys. Rev.* 60, 693.
Elliot, D. R. & King, L. D. P. 1941 *Phys. Rev.* 60, 489.
Ellis, C. D. 1922 *Proc. Roy. Soc.* A, 101, 1.
Ellis, C. D. 1938 *Proc. Phys. Soc.* 50, 213.
Feather, N. 1929 *Phys. Rev.* 34, 1558.
Feather, N. 1936 *Introduction to nuclear physics.* Cambridge.
Fünfer, E. 1939 *Ann. Phys., Lpz.*, 35, 147.
Gamow, G. 1931/1937 *Structure of atomic nuclei.*
Guggenheimer, K. M. 1940 *Nature, Lond.*, 145, 104.

Holmholz, A. C. 1941 *Phys. Rev.* **60**, 415.
Holloway, M. G. & Moore, B. L. 1940 *Phys. Rev.* **58**, 847.
Humphreys, R. F. & Pollard, E. 1941 *Phys. Rev.* **59**, 942.
Humphreys, R. F. & Watson, W. W. 1941 *Phys. Rev.* **60**, 542.
Jordan, P. 1937 *Ergebn. exakt. Naturw.* **16**, 47.
Krishnan & Nahum 1942 *Proc. Roy. Soc.* A, **180**, 321, 333.
Kuerti, G. & Wilkins, T. R. 1940 *Phys. Rev.* **57**, 1081.
Lewis, W. B. & Bowden, B. V. 1934 *Proc. Roy. Soc.* A, **145**, 235.
Margenau, H. 1941 *Phys. Rev.* **59**, 627.
May, A. N. & Vaidyanathan, R. 1936 *Proc. Roy. Soc.* A, **155**, 519.
McMillan, E. & Lawrence, E. O. 1935 *Phys. Rev.* **47**, 343.
Murrell, E. B. & Smith, C. L. 1939 *Proc. Roy. Soc.* A, **173**, 410.
Nordheim, L. W. & Yost, F. L. 1937 *Phys. Rev.* **51**, 942.
Peierls, R. 1940 *Reports on progress in physics*, **7**, 87.
Placzek, G. & Bethe, H. A. 1940 *Phys. Rev.* **57**, 1075.
Plain, G. P., Herb, R. G., Hudson, C. M. & Warren, R. E. 1940 *Phys. Rev.* **57**, 187.
Pollard, E. 1938 *Phys. Rev.* **54**, 411.
Pollard, E. 1939 *Phys. Rev.* **56**, 961.
Pollard, E. & Brasefield, C. J. 1936 *Phys. Rev.* **50**, 890.
Pollard, E., Davidson, W. & Schultz, H. L. 1940 *Phys. Rev.* **57**, 1117.
Pollard, E., Schultz, H. L. & Brubacker, G. 1938 *Phys. Rev.* **53**, 351.
Pollard, E. & Watson, W. W. 1940 *Phys. Rev.* **58**, 12.
Powell, C. F., May, A. N., Chadwick, J. & Pickavance, T. G. 1940 *Nature, Lond.*, **145**, 893.
Present, R. D. 1941 *Phys. Rev.* **60**, 28.
Rarita, W. & Schwinger, J. 1941 *Phys. Rev.* **59**, 436.
Richards, H. T. 1941 *Phys. Rev.* **59**, 796.
Roberts, A., Downing, J. R. & Deutsch, M. 1941 *Phys. Rev.* **60**, 544.
Rosenblum, S. 1937 *C.R. Acad. Sci., Paris*, **204**, 175, 345.
Rutherford, Lord E. 1929 *Proc. Roy. Soc.* A, **123**, 373.
Rutherford, Lord E. & Lewis, C. B. 1933 *Proc. Roy. Soc.* A, **142**, 347.
Schultz, H. L. & Watson, W. W. 1940 *Phys. Rev.* **58**, 1047.
Seaborg, G. T. 1940 *Chem. Rev.* **27**, 199.
Smith, C. L. & Murrell, E. B. 1939 *Proc. Camb. Phil. Soc.* **35**, 298.
Stephens, W. E. 1940 *Phys. Rev.* **57**, 938.
Stephens, W. E., Djanab, K. & Bonner, T. W. 1937 *Phys. Rev.* **52**, 1079.
Stuhlinger, E. 1939 *Z. Phys.* **114**, 185.
Szalay, A. 1939 *Z. Phys.* **112**, 29.
Teller, E. & Wheeler, J. A. 1938 *Phys. Rev.* **53**, 778.
Thibaud, J. & Comparat, P. 1939 *J. Phys. Radium*, **10**, 161.
Volkoff 1940 *Phys. Rev.* **57**, 866.
Wakatuki, T. 1940 *Proc. Phys.-Math. Soc. Japan*, **22**, 430.
Weisskopf, V. & Ewing, D. H. 1940 *Phys. Rev.* **57**, 472.
Weizsäcker, C. F. v. 1935 *Z. Phys.* **96**, 431.
Wigner, E. 1937 *Phys. Rev.* **51**, 947.
Wigner, E. 1939 *Phys. Rev.* **56**, 519.
Wigner, E. & Feenberg, E. 1941 *Reports on progress in physics*, **8**, 274.
Wilkins, T. R. & Kuerti, G. 1940 *Phys. Rev.* **57**, 1081.
Wilson, R. S. 1941 *Proc. Roy. Soc.* A, **177**, 382.

A new method of determining half-value periods from observations with a single Geiger counter

By A. G. WARD, *Emmanuel College, Cambridge**

A coincidence circuit is described which makes possible the determination of half-value periods between 10^{-4} sec. and 1 sec. using a single Geiger counter. This arrangement has been used to determine the half-value periods of actinium A, thorium A and radium C' with the following results: Ac A $(1\cdot83 \pm 0\cdot04) \times 10^{-3}$ sec., Th A $(1\cdot58 \pm 0\cdot08) \times 10^{-1}$ sec., Ra C' $(1\cdot48 \pm 0\cdot06) \times 10^{-4}$ sec. A discussion of the limitations of the method and of possible sources of error is given.

INTRODUCTION

The method of determining half-value periods described in this paper is based on the analysis of the distribution in time of the impulses recorded in a single Geiger counter. This analysis is made possible by the use of an electrical circuit with a variable resolving time, developed for this work by Dr J. V. Dunworth in the Cavendish Laboratory.† The half-value periods of actinium A, thorium A and radium C' have been measured by this method. The values obtained are as follows:

$$\text{Ac A } (1\cdot83 \pm 0\cdot04) \times 10^{-3} \text{ sec.}, \quad \text{Th A } (1\cdot58 \pm 0\cdot08) \times 10^{-1} \text{ sec.},$$

$$\text{Ra C' } (1\cdot48 \pm 0\cdot06) \times 10^{-4} \text{ sec.}$$

The periods of thorium A and actinium A were determined by Moseley & Fajans (1911) using a rotating disk method. The half-value period of radium C' has been determined by Rotblat (1939, 1941) (see also Dunworth 1939). It will be seen that the half-value periods reported in this paper agree reasonably well with previous values. Although, up to the present time, the experimental technique now to be described has been tested only in the determination of half-value periods of elements of the naturally radioactive series, it has many other applications in the field of nuclear physics.‡ It will clearly be of use in the search for new radio-

---

* *Editorial note.* The first draft of this paper was prepared by Mr Ward, to whom all the experimental observations are due. Mr Ward's present duties have, however, made it impossible for him to devote further time to the final revision of the paper, and this I have undertaken on his behalf. I also took the opportunity of quoting Mr Ward's final results in a recent review article of my own (Feather 1941).—N. FEATHER.

† The present experiment was started in the Cavendish Laboratory by Mr T. B. Scheffler in January 1939 but had to be abandoned by him before any real success had been attained. The experimental work carried out by the writer was completed in April 1940.

‡ Since the work described in this paper was completed, two papers by Driscoll, Hodge & Ruark (1940) and Roberts (1941) have been published which describe circuits having the same principles of operation as that employed by the writer. The circuit described by Roberts, in particular, very closely resembles the one described in this paper, but the writer feels that, in spite of this, the present circuit has merits which justify its description in detail. In the papers quoted the new circuits are discussed in relation to the study of the behaviour of Geiger counters, but Roberts (1941) also refers to possible applications to the measurement of half-value periods of radioactive substances.

activities with half-value periods of this order produced by the decay of longer-lived parent substances, and it should also prove useful in the investigation of γ-ray transitions of comparatively long lifetime (1 sec. $> \tau > 10^{-4}$ sec.).

METHOD OF ANALYSING COINCIDENCE COUNTING RATES

The electrical circuit, which is described more fully later in the paper, has the property of recording a 'coincidence' whenever any impulse in the Geiger counter occurs within a time interval T of the preceding impulse. It can also be used to count the total number of impulses occurring in the counter in a given time.

When the impulses produced in the Geiger counter are randomly distributed in time, and the electrical resolving time of the associated circuit is T, it can be shown that the coincidence counting rate Co is given by the equation

$$Co = n(1 - e^{-nT}), \qquad (i)$$

where n is the average number of impulses per unit time. The coincidence counting rate is here understood as the average number of times per second that an impulse follows the last preceding impulse within a time interval less than T sec.

Suppose now that a small amount of actinon gas, for example, is introduced into such a counter. Then, because successive disintegrations occur, the impulses will no longer be distributed at random. The total counting rate will be the sum of the 'natural effect' of the counter and the effect produced by the decay of the actinon, actinium A and the succeeding members of this radioactive series. If all the radioactive atoms disintegrating in the counter are recorded, and if the average time interval between disintegrations is large compared with the half-value period of actinium A (the shortest-lived member of the series), then it is evident that each impulse due to the disintegration of an atom of actinon will be followed by an impulse due to the disintegration of the daughter atom actinium A, and that the separation in time of such a pair of impulses will be much less than the average time interval between recorded impulses in general. The coincidence rate C can now be shown to be

$$C = (n_a + n_b)\,[\,1 - e^{-(n_a+n_b)\,T - n_b(1-e^{-\lambda T})/\lambda}\,] + n_b[\,1 - e^{-(n_a+n_b+\lambda)\,T - n_b(1-e^{-\lambda T})/\lambda}\,], \qquad (ii)$$

where n_a is the number of 'random' impulses per unit time and n_b is the number of pairs of correlated impulses per unit time, the correlation being expressed by the probability $P_T = (1 - e^{-\lambda T})$ that the second impulse of a pair follows the first within a time T. Here λ is the characteristic radioactive decay constant concerned. It should be noted that the total counting rate in this case is $n_a + 2n_b$. The precise significance of the counting rates n_a and n_b in our present example has now to be discussed.

In the first place the correlation in time between the impulses produced by the decay of actinium B and successive members of the radioactive series is not considered in this analysis. Since the half-value periods of these substances are very much longer than the resolving times T of interest for the determination of

the period of actinium A, they are considered as contributing 'random' counts, included, with the natural effect of the counter, in the value of n_a. Again, there are two reasons why some of the 'correlated' disintegrations of actinon and actinium A contribute to n_a rather than to n_b. The first reason is that the efficiency of the Geiger counter is not exactly 100 %, the second that there is a finite time τ in which the counter is insensitive to a further impulse after each impulse which it records. If ϵ is the efficiency of the counter, only a fraction ϵ of the disintegrations of actinon are counted, and a similar fraction of the disintegrations of actinium A would be counted, if the second effect were negligible. If this were the case, a fraction ϵ^2 of the correlated disintegration pairs $An \rightarrow AcA \rightarrow AcB$ would be recorded, and a fraction $\epsilon(1 - \epsilon)$ of the disintegrations of each body would give rise to uncorrelated impulses counted in n_a. If the finite recovery time τ is not negligible, as we have just assumed, a further fraction $\epsilon^2(1 - e^{-\lambda\tau})$ of the disintegrations of actinium A fail to be recorded, only $\epsilon^2 e^{-\lambda\tau}$ of the total number of 'correlated' pairs go to make up the observed rate n_b, and an additional fraction $\epsilon^2(1 - e^{-\lambda\tau})$ of the disintegrations of actinon—over and above the fraction $\epsilon(1 - \epsilon)$ previously transferred—must be regarded as contributing to n_a. The effect of the finite recovery time is not, of course, confined to the correlated disintegrations; its general effect on the coincidence rate can be represented by interpreting T in equation (ii) as a time measured from a zero τ sec. after the occurrence of each impulse recorded—after the redistribution already made between n_b and n_a.

For the purposes of discussion the actinon series has been used to illustrate the application of equation (ii) in the analysis of coincidence observations. It will be realized that similar, but in each case distinct, considerations hold when equation (ii) is used in the determination of the half-value periods of thorium A and radium C'.

In general, for the determination of the period of a radioactive substance, the following procedure is carried out. Values of C are obtained for different known values of T. The single counting rate $n_a + 2n_b$ is also measured. From these measurements, with the help of equation (ii), λ is determined. The half-value period is then $0 \cdot 693/\lambda$. From an inspection of equation (ii) it can be seen that a determination of λ is simplest when λ is greater than $(n_a + n_b)$, and when n_a is kept as small as possible.

DESIGN OF GEIGER COUNTERS AND OF THE COINCIDENT CIRCUIT

The Geiger counters used in these experiments were made of small dimensions. This was necessary to keep the value of n_a (equation (ii)) as small as possible, since a small counter has a small natural effect and, at the counting rates used, the natural effect contributes appreciably to n_a. A small counter is also relatively insensitive to the γ-rays from radioactive sources in the vicinity. In addition, some evidence was obtained in the course of these experiments which showed that small counters have a shorter recovery time (τ) than larger counters of the same construction.

The counters were of the type recommended by Curran & Petrzilka (1939), having a brass cylinder and a 50μ tungsten wire, and were filled with a mixture of 90 % argon and 10 % alcohol to a pressure of 12 cm. of mercury. Figure 1a shows the detailed construction of the counters. The brass cylinder is a sliding fit in the pyrex glass tube and glass end-pieces keep the tungsten wire central and enclose the counting volume. The method of attaching a spring to keep the central wire taut is shown in the figure. Wire leads to the counter case and the central wire come through constrictions in the glass tube and vacuum seals are made with solder and sealing wax.
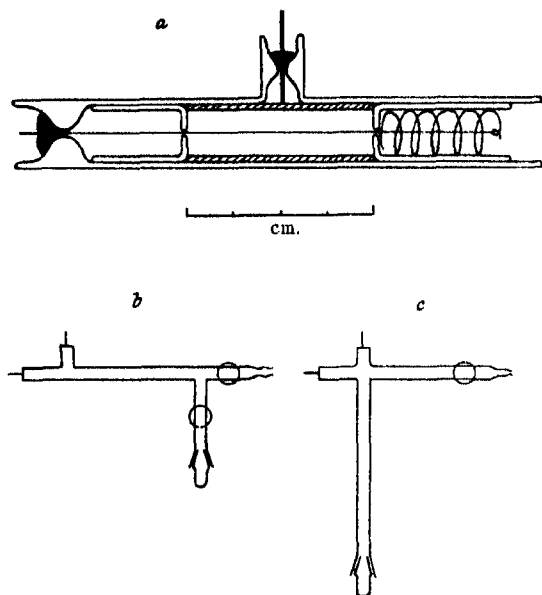
FIGURE 1

Two different methods were used to introduce the radioactive material into the counting volume. Figure 1b illustrates the arrangement used to introduce radon and thoron into the counters. The emanating source was placed in a receptacle fitting into the side tube with a ground-glass joint, and the entrance of the radio-active gas into the part of the tube containing the Geiger counter was controlled by a tap. When the tap was open, the radioactive gas diffused into the counter proper through the small hole in the glass end-piece used to centre the tungsten wire.

Figure 1c shows the arrangement used in measuring the half-value period of actinium A. The emanating source was placed in a cylindrical container which was introduced into the long side tube through a ground-glass joint. A small piece of soft iron was attached to the source container, and enabled it to be moved along the tube by means of an external magnet. When the source container was at the end of the side tube farthest from the counter, the actinon gas atoms decayed

before they had time to diffuse along the tube as far as the counter. With the source moved closer to the counter, the actinon atoms were able to diffuse into the counter through a small hole drilled in the brass wall of the counter. This hole did not face directly down the side tube but was arranged in such a way that particles arising from disintegrations in the side tube would have small chance of passing through the hole into the counting volume. It was shown experimentally that the number of actinon atoms diffusing into the counter was reduced to about 50 % by moving the source 1 cm. farther away. This is in reasonable agreement with the value calculated from the rate of diffusion of the actinon atoms and their known half-life (3·9 sec.).
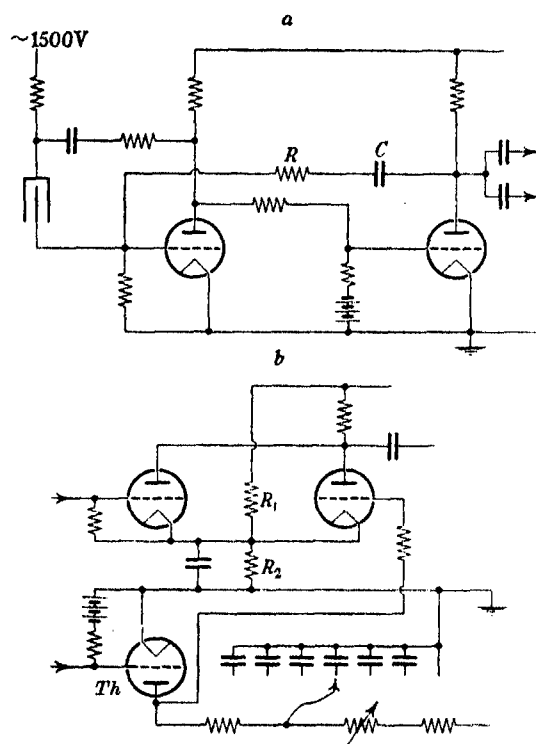
FIGURE 2

The electrical circuit used for the experiments described in this paper is shown in figure 2. For the sake of simplicity the valves are shown as triodes. Figure 2*a* shows a multivibrator circuit and Geiger counter which operates in a manner suggested by Getting (1938). This circuit is very sensitive to small voltage changes on the grid of the first valve, if the bias of the second valve is correctly adjusted. Two purposes are served by the use of this circuit. The voltage pulse on the plate of the second valve is almost independent of the size and length of the pulse of applied voltage, and the changes in voltage on the plate and grid of the first valve can be utilized to provide a quenching action for the counter. The arrangements

shown for quenching the counter are similar to those suggested by Getting (1938) except that a larger range of useful operating voltage is achieved by applying quenching voltage both to the case and to the wire of the Geiger counter.

A single impulse in the Geiger counter produces a rectangular pulse of negative voltage at the plate of the second valve. The length of this pulse (see figure 3a) can be altered by varying the values of C and R (figure 2a). Pulse lengths between 10^{-3} and 10^{-4} sec. have been used in these experiments, the values being measured by visual observations on an oscillograph. These are the values of the time interval, τ, for which the Geiger counter is rendered useless for counting by the quenching voltage.
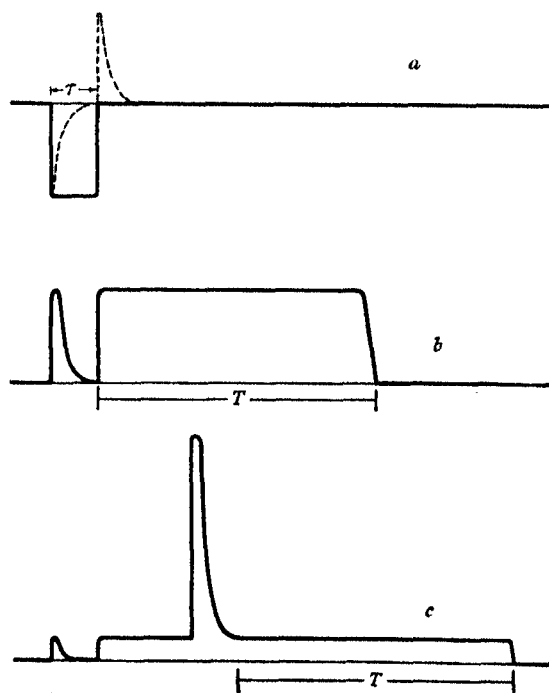
FIGURE 3

It was not possible to use pulse lengths shorter than 10^{-4} sec., since then a single impulse in the Geiger counter would trip the multivibrator circuit more than once.

The rectangular pulses of negative voltage from the second valve in figure 2a are fed into the second part of the electrical circuit, figure 2b, by two separate paths. The resistance-capacity coupling used has a small time constant and alters the rectangular pulse to the shape shown by the dotted line in figure 3a.

The two upper valves in figure 2b have a common plate resistance and form a modified Rossi coincidence pair. The cathode potential of this pair of valves is maintained at about 100 V positive with respect to ground by the potential dividers R_1, R_2. The control grid of the second valve of this pair is connected

through a megohm resistance to the plate of the thyratron (*Th*). The voltage pulse represented by the dotted line, figure 3*a*, is fed to the grid of the thyratron and the grid of the first valve of the Rossi circuit. The voltage changes occurring across the plate resistance of the Rossi circuit are shown in figure 3*b*. The initial short pulse is caused by the first valve of the Rossi pair being rendered non-conducting by the negative part of the input pulse. The second positive part of this input pulse strikes the thyratron and the consequent fall of the plate potential below 100 V is communicated to the grid of the second valve of the Rossi pair and this valve is made non-conducting. This valve remains non-conducting for a length of time T determined by the time necessary to charge the plate of the thyratron up to 100 V, and this time T is determined by the value of the resistance and capacity in the plate circuit of the thyratron.

The output voltage changes occurring at the plates of the Rossi pair when two impulses are recorded within a time T are shown in figure 3*c*. The second impulse is recorded as a coincidence by the large voltage change which occurs at the plates of the Rossi pair. It should be noted that the second valve of the Rossi circuit is again held non-conducting for a length of time T after the second impulse, and the circuit will again record a coincidence if another impulse falls within this time. Thus the electrical circuit fulfils the condition stated previously: it records a coincidence every time an impulse follows the preceding impulse within a time T.

The output from the plates of the Rossi pair is fed through a discriminator valve into a scale of eight and mechanical recorder. The resolving time of the scale of eight circuit was shown experimentally to be less than 10^{-4} sec.; thus the correction for counting losses was dependent only on the recovery time of the mechanical recorder and on that of the Geiger counter and its associated multivibrator circuit. Corrections for loss in the mechanical circuit were avoided by using a scale of sixty-four when fast counting rates were required. With a recovery time of 10^{-4} sec. for the Geiger counter, losses in the counting of randomly occurring impulses for this reason were always less than 1 %.

Two methods were used to measure the resolving times T of the coincidence circuit obtained with different settings of the variable resistance and capacities of figure 2*b*. For resolving times between 1·0 and 0·02 sec. a neon stabilizer was used as a pulse generator and synchronized with the 50 cycle a.c. By this means pulses were obtained with a frequency of 50 per sec. or any submultiple of 50 per sec. The resolving time T equal to the time between succeeding pulses could then be easily determined by finding the values of capacity and resistance in the plate circuit of the thyratron for which the circuit recorded each pulse as a coincidence. For resolving times less than 0·02 sec. the resolving time was determined by measuring the chance coincidence rate for a known single counting rate when the impulses had a random distribution in time.

By observing the coincidence rate at various single counting rates for a fixed value of T (plate resistance and capacity constant) the validity of equation (i) was proved. This test also showed that the coincidence circuit was behaving correctly and that no appreciable number of 'spurious' coincidences was being recorded.

Such 'spurious' coincidences were found when the pulse length of the multivibrator was made less than 10^{-4} sec.

It may be pointed out that the electrical circuit which has just been described can also be used for investigations of the behaviour of a Geiger counter (cf. Roberts 1941). It would be of interest to test one of the fast counters described by Neher (1938) with this coincidence circuit. These counters are said to have a recovery time of 10^{-6} sec. If it is possible to use a multivibrator pulse length of this order of magnitude the range of half-value periods which could be investigated by the present method would be greatly increased. The writer feels that the possibility of utilizing resolving times of this order is rather remote, since it would appear probable that the collection time of ions in a counter is of the order of 10^{-5} sec.

EXPERIMENTAL RESULTS AND THEIR ANALYSIS

The following procedure was used to determine the half-value period of actinium A. The natural effect of the counter was first measured with the source at the end of the side tube (figure 2c) remote from the counter. The emanating source was then moved to a predetermined position close to the counter, so chosen that a suitable counting rate was obtained from the emanation, in diffusion-decay equilibrium within the counter. Alternate readings were then taken of the total counting rate (which increased with time due to the growth of actinium B together with succeeding members of the actinium series) and of the 'coincidence' rate (at various resolving times). At the end of this sequence of observations the emanating source was again moved back to the end of the side tube. Since actinon itself has a short half-value period (3·9 sec.), it disappears rapidly when the source is removed; after a short time, therefore, the observed counting rate was due only to actinium B and the succeeding members of the series and to the natural effect as previously determined. When these experiments were made the equipment available did not allow of simultaneous determinations of the total counting rate and the coincidence rate; clearly if this had been possible it would have been preferable to the procedure of alternate determinations actually adopted.

Tables 1 and 2 contain the analysis of one set of observations used in the determination of the half-value period of actinium A. The time in minutes recorded in column I (tables 1 and 2) is the time which had elapsed since the emanating source was brought near to the counter. Column II, table 1, gives the total counting rate (in min.$^{-1}$) observed during the 229 min. for which the source remained in position, as well as the corresponding rate for the next 21 min. after the source had been removed. In column V the natural effect (54 min.$^{-1}$) has been subtracted from these latter rates and column VI represents their correction to the time of removal of the source ($t = 229$ min.). We take the average of these corrected rates (709 ± 6 min.$^{-1}$) as the steady equilibrium counting rate due to actinium B and subsequent products (equilibrium can be regarded as virtually complete after 229 min.), and from this value we obtain the counting rate due to these products at any time during their growth to equilibrium (column III).

TABLE 1

(For explanation see text)

I	II	III	IV	I	II	V	VI
10	750 ± 10	85	665	229	Actinium source removed		
28	951 ± 11	257	694	231	748 ± 19	694	707
45	1081 ± 12	384	697	233	768 ± 20	714	740
63	1173 ± 12	479	694	235	740 ± 19	686	729
83	1215 ± 12	553	662	238	684 ± 13	630	702
114	1329 ± 13	625	704	242	646 ± 13	592	705
157	1346 ± 13	676	670	246	588 ± 12	534	683
184	1387 ± 13	691	696	250	562 ± 12	508	696
222	1375 ± 11	701	674				

TABLE 2

(For explanation see text)

I	II	III	IV	V	VI	VII
19	340 ± 6·5	10^{-2}	861	59	280	339
37	280 ± 6	4×10^{-3}	1016	41·6	226	268
54	183·4 ± 4·8	2×10^{-3}	1122	28·4	155·5	184
73	139·2 ± 3·7	$1·41 \times 10^{-3}$	1187	23·4	122·4	145·8
92	452 ± 7	10^{-2}	1267	158	280	438
103	85·4 ± 0·8	7×10^{-4}	1291	14·6	69·6	84·2
127	57·1 ± 2	$4·4 \times 10^{-4}$	1332	10·0	45·6	55·6
144	26·1 ± 1·3	$2·2 \times 10^{-4}$	1351	5·3	24·0	29·3
193	459 ± 8	10^{-2}	1380	194	280	474
202	312 ± 6	4×10^{-3}	1384	88	226	314
211	54·4 ± 2·6	$4·4 \times 10^{-4}$	1387	11·0	45·8	56·8

Column IV, which gives the difference between the entries in columns II and III, is then the calculated counting rate due to actinon and actinium A, together with the natural effect and the additional γ-ray background from the emanating source in its near position. It is some confirmation of the basis of our calculations that the values in column IV are approximately constant, showing no regular trend as t increases, and we take the mean (684 ± 5 min.$^{-1}$) as the effective counting rate due to actinon, actinium A and the background effect over this part of the experiment. Table 2 contains the coincidence counting rates together with their analysis. Columns II and III give the observed coincidence rates (min.$^{-1}$) and the corresponding resolving times (sec.). Values of the total counting rate $(n_a + 2n_b)$, interpolated with the aid of table 1, are given in column IV. The analysis has been carried out as follows: for the largest value of the resolving time (in this case $T = 0·01$ sec.), λT may be regarded as large and $e^{-\lambda T}$ as negligibly small, to a first approximation. Then equation (ii) becomes

$$C = (n_a + n_b)\,[1 - e^{-(n_a + n_b)T}] + n_b,$$

and a knowledge of the coincidence counting rate, C, and of the total rate $(n_a + 2n_b)$ allows first values of n_a and n_b to be calculated. These values are then used to obtain values of λ from the coincidence observations at shorter resolving times,

using equation (ii) in its complete form. A better approximation to values of n_a and n_b may then be obtained from the observations at large resolving times, and a better approximation to λ by the use of the corrected values of n_a and n_b. Eventually, after successive approximations, values of λ, n_a and n_b are obtained giving good agreement with all the observations. With the results given in tables 1 and 2 we obtain as best values, $n_b = 285$ min.$^{-1}$ and $\lambda = 3\cdot78 \times 10^2$ sec.$^{-1}$. The extent of the agreement reached with these values of the constants is shown in columns V–VII, table 2. Column V gives the calculated values of

$$(n_a + n_b) \,[\, 1 - e^{-(n_a+n_b)\, T - n_b(1-e^{-\lambda T})/\lambda}\,],$$

and column VI the corresponding values of

$$n_b \,[\, 1 - e^{-(n_a+n_b+\lambda)\, T - n_b(1-e^{-\lambda T})/\lambda}\,].$$

The sum of these two values, given in column VII, is the calculated coincidence rate, to be compared with the observed rate of column II. The contributions of columns V and VI to the calculated rate may be thought of as the contributions of the chance and true coincidences, respectively. It will be seen that the agreement is reasonable and that chance coincidences represent only a small fraction of the observed coincidence rate in most cases. From the results of tables 1 and 2, and of two similar experiments, the final value for the half-value period of actinium A was determined as $(1\cdot83 \pm 0\cdot04) \times 10^{-3}$ sec. The probable error in this value is in part due to uncertainties of the order of 1 % in the coincidence resolving times, T.

It is of some interest to compare the calculated value of n_b with the total counting rate due to the disintegrations of actinon and actinium A recorded in the counter. The single counting rate observed in the presence of the source, excluding the effect produced by actinium B and subsequent products, is 630 (684 − 54) min.$^{-1}$ (table 1, column IV). But $2n_b$, the rate due to the paired ('coincident') disintegration of actinon and actinium A, is $2 \times 285 = 570$ min.$^{-1}$. At this rate of counting, on the average, actinon-actinium A disintegration pairs of less than 10^{-4} sec. separation occur 11 times per min. (10^{-4} sec. was the time for which the counter was kept out of action by the quenching voltage), and thus only 49 single counts per minute (630 − 570 − 11) remain unexplained in this analysis. Some of these must certainly have been caused by the decay of actinon and actinium A atoms outside the counter—the α-particles from these disintegrations penetrating the counting volume through the hole allowing entrance to the actinon gas—and there is the additional γ-ray effect due to the emanating source in the near position. It is easy to believe that in one or other way the whole discrepancy can be explained; if so we must conclude that practically 100 % of the atoms of actinon and actinium A disintegrating inside the counter are recorded. Some estimate of the efficiency of the counter in recording β disintegrations can now be made from the contribution to the total counting rate assigned to actinium B and subsequent products in equilibrium ($t = 229$ min., table 1). This equilibrium counting rate of 709 min.$^{-1}$ is to be ascribed to the disintegration of the active deposit in equilibrium with an amount of emanation

producing 296 (285 + 11) disintegrations per min. Leaving out of account the branching at actinium C, and neglecting the difference between secular and transient equilibrium, we thus have a counting rate of 413 min.$^{-1}$ (709 − 296) representing the 592 β disintegrations per minute due to actinium B and actinium C″, if 100 % efficiency be assumed for the counting of the α disintegrations of actinium C. Thus an efficiency of about 70 % for the registration of the β disintegrations of atoms situated on the negatively charged counter wall is to be deduced from the observations (the active deposit will form on the counter wall as in the normal method of collecting the deposit by the method of recoil).

For the determinations of the half-value periods of thorium A and radium C′ the counter construction shown in figure 2b was used. An old preparation of radiothorium containing a small amount of radium was employed as the emanating source. In the determination of the half-value period of thorium A the strength of this source was adjusted to give about 50 impulses per min. due to the $Tn \xrightarrow{\alpha} ThA \xrightarrow{\alpha}$ disintegrations taking place inside the counter. The 'contamination' due to the presence of radium in the source then amounted to about 40 impulses per min. if the radon was allowed to grow to equilibrium. By periodically pumping out and refilling the counter, however, observations were made with a much smaller 'contamination' effect than this. The 'natural' effect of the counter was about 8 impulses per min. with this arrangement. For these measurements with thorium A the time for which the counter was kept out of action by the multi-vibrator was increased to about 10^{-3} sec.: this had the advantage of largely suppressing unwanted coincidences due to the $RaC \xrightarrow{\beta} RaC' \xrightarrow{\alpha}$ disintegrations arising from the contamination, whilst not seriously affecting the main determination, because of the generally small counting rates which had necessarily to be employed with this relatively long-lived body (half-value period $\sim 10^{-1}$ sec.). In order to obtain reasonable statistical accuracy at these small rates, observations were extended over a total period of about 2 weeks. After final analysis of the results (such a scattered body of data obviously cannot be presented here in concise form), the half-value period of thorium A was calculated as $(1 \cdot 58 \pm 0 \cdot 08) \times 10^{-1}$ sec.

The determination of the half-value period of radium C′ was carried out in the following way. A much larger amount of the same emanating material was placed in the side tube of the arrangement of figure 2b and the radon was allowed to accumulate over the preparation for some days. Then the tap communicating with the rest of the apparatus was opened for about 30 sec., allowing some radon to diffuse into the counting volume. A relatively large amount of thoron was admitted to the counter by this procedure, but it rapidly decayed (half-value period 54 sec.) leaving effectively pure radon. This was left in the counter overnight, by which time the residual activity due to thorium active deposit had further diminished in relation to the activity due to the corresponding members of the radium series, and observations were then taken of the total counting rate and of the coincidence rate corresponding to various resolving times. Because, in this case, the radio-element of short life under investigation is not the first member of the active deposit series, but the fourth, a much larger ratio of total counting rate

to true coincidence rate was necessary than with thorium A and actinium A. Also, the true coincidences were now those between a β-particle (that of radium C) and an α-particle (from radium C′), rather than between two α-particles. For both reasons the general efficiency of the coincidence method was less than before.

A second set of experiments leading to the half-value period of radium C′ was therefore carried out as follows. The radon was introduced into the counter, as previously described, but it was pumped out again, and the counter flushed and refilled, as soon as the active deposit had grown to equilibrium. Observations were then taken of the total counting rate, and of the coincidence rate at various resolving times, as the active deposit decayed. It can be seen that the periods of successive members of the radium series are such that the contribution of the RaC → RaC′ → disintegrations to the total counting rate is appreciably greater in this case than in the last, provided, of course, that the natural effect of the counter is small compared with the total counting rate employed. In spite of the added complication of the need to calculate the amount of radium C present at any time after the beginning of the experiment,* this type of experiment, using a decaying rather than an effectively steady source, gave the best results for the half-value period of radium C′. A representative set of results, belonging to one experiment with a decaying source, is given, with explanatory notes in table 3. The final value for the half-value period of radium C′ deduced from several such experiments was $(1.48 \pm 0.06) \times 10^{-4}$ sec. In these experiments the time during which the counter was extinguished was about 10^{-4} sec. Since this time is of the same order as the half-value period of the body under investigation, and was not known accurately, it is not possible to deduce, from observations similar to those recorded in table 3, any precise information concerning the efficiency of the counter in respect of α and β disintegrations taking place within it, as was done in the earlier experiment on actinium A.

DISCUSSION

From the examples studied, it appears that the method of measuring half-value periods described in this paper should be capable of good results over the range of periods from 10^{-4} to 1 sec.

The shortest period which can be measured is limited by the recovery time of the Geiger counter; in the present experiments recovery times shorter than 10^{-4} sec. were not achieved. It is probable that the nature of the discharge process is such that a natural lower limit will be set at about this point.

The longest period which can be measured is limited by the natural effect of the counter. The natural effect is a function of counter size and of the cleanliness of the constructional materials used. Counters have been reported having 'naturals' of no more than 3 or 4 impulses per min.: the counter used in the experiments

---

* A further complication, though not a serious one, arose from the fact that it was never possible to remove more than about 99 % of the radon by simple pumping. A small correction had thus to be made for the residue.

TABLE 3

I	II	III	IV	V	VI	VII	VIII	IX
11·5	170·7 ± 1							
16	152·1 ± 1							
21		140·5	24·4 ± 0·3	4×10^{-4}	21·7	6·02	18·5	24·5
25	132·2 ± 1							
30		122·2	8·86 ± 0·16	10^{-4}	19·6	1·21	7·48	8·69
35	112·7 ± 1							
40		103·3	9·68 ± 0·17	$1·5 \times 10^{-4}$	17·05	1·27	8·73	10·0
46	92·4 ± 0·6							
52		83·3	19·20 ± 0·23	10^{-3}	14·0	4·86	13·9	18·8
58	74·5 ± 0·6							
65		65·4	10·19 ± 0·15	4×10^{-4}	11·05	1·28	9·33	10·6
72	56·9 ± 0·5							
79		49·0	3·52 ± 0·09	10^{-4}	8·38	0·19	3·16	3·35
86	42·6 ± 0·4							
92		37·8	7·30 ± 0·14	10^{-3}	6·39	1·02	6·33	7·35
98	33·3 ± 0·4							
105		29·0	2·96 ± 0·08	2×10^{-4}	4·80	0·13	2·92	3·05
112	25·5 ± 0·3							
208	5·23 ± 0·15							

$n_b = 24·0$ sec.$^{-1}$ at zero time. $\lambda = 4680$ sec.$^{-1}$.

Column I: Time in minutes from removal of radon from counter.
Column II: Total counting rate (sec.$^{-1}$).
Column III: Total counting rate $(n_a + 2n_b)$ by interpolation from column II.
Column IV: Observed coincidence rate (sec.$^{-1}$).
Column V: Resolving time T (sec.).
Column VI: Calculated value n_b (allowance for decay from zero time) (sec.$^{-1}$).
Column VII: $(n_a + n_b)\,[1 - e^{-(n_a+n_b)\,T} - n_b(1-e^{-\lambda T})/\lambda\,]$.
Column VIII: $n_b\,[1 - e^{-(n_a+n_b+\lambda)\,T} - n_b(1-e^{-\lambda T})/\lambda\,]$.
Column IX: Calculated coincidence rate (sec.$^{-1}$). Column VII + column VIII.

on thorium A and radium C′ had a natural effect of 8 per min. initially, but the introduction of radon resulted in a small increase of this rate, due to the residual activity of the radium D, E and F which remained after the short-lived bodies had decayed. Even with half-value periods of $\frac{1}{10}$ sec. long periods of counting are necessary, particularly for the determination of coincidence rates at small resolving times, and this represents another aspect of the limiting factor. It does not appear likely that the coincidence method using a single counter will be at all satisfactory for half-value periods greater than 1 sec.

Returning to the consideration of the use of the method for the shortest half-value periods capable of study, it must be admitted that there may well be several sources of inaccuracy in the final determinations. When the half-value period is of the same order of magnitude as the pulse length, it is evident that many true coincidences will be missed—the second of a pair of correlated disintegrations occurring whilst the multivibrator circuit is still tripped and the counter quenched. This would be of minor importance, since it could be allowed for completely, if entirely dependable quenching could always be obtained, but, particularly when α-particles are in question, it is doubtful whether this is the case. Since the

starting voltage of the counter is smaller for α-particles than for β-particles, it is difficult to arrange for conditions of quenching and recovery which are equally good for the two kinds of particle. If the counter is not completely quenched whilst the multivibrator circuit is tripped, it may continue to discharge when the full voltage is restored, and a 'spurious' coincidence may result. When the circuit was tested for chance coincidences, using a γ-ray source, no evidence of such spurious coincidences was obtained until the counter voltage was raised considerably above the starting voltage for β-particle counting. In the experiments with radium C', already described, the voltage applied to the counter was just sufficient to bring the operating point on to the 'plateau' (for β counting). The close agreement of the results of these experiments with those which Dunworth and Rotblat obtained, under conditions in which no similar errors could arise, seems to indicate that this precaution had the desired effect of reducing these errors, in the present experiments, to negligible proportions. Since the uncertainty on this point remains, the limits of error here set in the radium C' determination are wider than the limits calculated from a purely statistical treatment of the observations.

A final point of discussion of present results concerns the efficiencies which have been deduced relative to the recording of α and β disintegrations taking place within the counters. We have to explain why 100 % of the α disintegrations and 70 % of all the β disintegrations are recorded. There is no difficulty in this explanation in the former case. For atoms situated on the counter wire (or on the wall) it is almost certain that either the α-particle or the recoil atom will traverse the effective volume of the counter—and the number of ion-pairs produced in the small counter will not be very different in the two events. Also, an α-particle entering the wire (or the wall) will probably produce a number of secondary electrons which will make detection more certain. In the case of β disintegrations the position is not so clear-cut, but it is again obvious from the experiments that a considerable proportion of disintegrations resulting in the projection of the β-particle into the metal support are registered, along with those disintegrations in which the β-particle is emitted directly into the counter. It seems unlikely that conditions are favourable to an efficiency of β recoil high enough to explain the observations, but when the production of secondary electrons, and 'reflexion' of the original β-particle, are taken into account there is nothing essentially surprising in the 70 % efficiency of recording actually found.

REFERENCES

Curran & Petrzilka 1939 *Proc. Camb. Phil. Soc.* **35**, 309.
Driscoll, Hodge & Ruark 1940 *Rev. Sci. Instrum.* **11**, 241.
Dunworth 1939 *Nature, Lond.,* **144**, 152.
Feather 1941 *Reports on progress in physics,* **7**, 66.
Getting 1938 *Phys. Rev.* **53**, 103.
Moseley & Fajans 1911 *Phil. Mag.* **22**, 629.
Neher 1938 *Procedures in experimental physics,* p. 266. New York: Prentice Hall.
Roberts 1941 *Rev. Sci. Instrum.* **12**, 71.
Rotblat 1939 *Nature, Lond.,* **144**, 248.
Rotblat 1941 *Proc. Roy. Soc.* A, **177**, 260.

Flame spectra in the photographic infra-red

By A. G. Gaydon

Leverhulme Research Fellow, Chemical Technology Department,
Imperial College, London, S.W.7

[Plates 6, 7]

The spectra of the flames of hydrogen, methane and carbon monoxide burning with oxygen and with nitrous oxide have been photographed in the region 6000–10,000 A. All flames in which water is a final product show a system of emission bands from the red to the far infra-red, the bands increasing in strength to longer wave-lengths. Outstanding heads have been observed at $\lambda\lambda 6165$, 6457, 6919, 7164, 8097, 8916, 9277 and 9669. It is shown that these bands are due to the vibration-rotation spectrum of H_2O. The top of a flame of oxygen burning in hydrogen is coloured red by the emission of these bands. In the hydrogen flame the bands are probably excited mainly thermally, but the strength of these same H_2O bands in the flame of moist carbon monoxide indicates that in this flame the excitation is a result of the combustion processes; this agrees with earlier theories on the formation of vibrationally activated molecules of CO_2 in this flame. In the hydrogen—nitrous-oxide flame new band structure in the infra-red is provisionally assigned to an extension of the ammonia α band. The methane—nitrous-oxide flame also shows the ammonia α band, and in addition strong emission of the red system of CN.

INTRODUCTION

The spectra of most ordinary flames have been investigated, at least qualitatively, in the visible and near ultra-violet regions of the spectrum, that is, roughly from 2000 to 7000 A. The spectra of the more common flames (hydrogen, carbon monoxide, methane) have also been studied in the far infra-red beyond 1μ (10,000 A), but the region between 7000 and 10,000 A appears to have been almost entirely neglected. This is, no doubt, largely because of the fact that flames are regarded as relatively feeble sources of light, and plates sensitized to the infra-red are relatively slow.

It has been shown by Kitagawa (1936) that the flame of oxygen burning in hydrogen emits some faint bands in the visible region, principally in the red. These bands were tentatively attributed to the vibration-rotation spectrum of the water

molecule. This assignment of the bands is shown, by the present work, to be correct, but did not at the time appear very convincing to the author. If the bands, which were also observed by the author in an ordinary oxy-hydrogen flame, were really due to the vibration-rotation spectrum of water, then it seemed that additional stronger bands should be observed in the photographic infra-red. Fortunately a large aperture spectrograph giving reasonable dispersion with high light-gathering power was available, and so it was decided to attempt a study of the spectrum of the oxy-hydrogen flame in the near infra-red with a view to checking whether the bands reported by Kitagawa in the visible were connected with the known vibration-rotation spectrum of water in the far infra-red.

It was at once found that the emission from the oxy-hydrogen flame was unexpectedly strong in the photographic infra-red, and satisfactory plates showing new band structure could be obtained with exposures of only a few minutes even down to 10,000 A. The investigation of a number of other flame spectra, including flames maintained by nitrous oxide, was then undertaken, and the results are presented here. The spectra of the various flames used were also photographed in the visible region, and some new features of flames maintained by nitrous oxide are also pointed out.

EXPERIMENTAL

Most of the flames whose spectra are described here were of pre-mixed gases burning at an ordinary blow-pipe jet; it may be assumed that all flames were of this type unless stated to the contrary. For flames of one gas in another (e.g. oxygen burning in hydrogen) a burner consisting of two vertical concentric quartz tubes of internal diameters 3 and 30 mm. was used. The base of the larger tube was closed by a metal fitting, through the centre of which the smaller tube could be slid, and through which three holes were drilled for the entry of the gas supporting the combustion. When the supporting gas was itself inflammable (e.g. hydrogen) the excess was burnt at the top of the larger tube; when the supporting gas was not inflammable a steel cap with a small hole through it was fitted over the top of the outer tube to reduce back draughts of air.

The large aperture glass prism spectrograph was used for most of the work. This instrument has a focal ratio of about 1 : 4 and the dispersion, in the adjustment used, which was slightly off minimum deviation to increase the dispersion, varied from 90 A/mm. at 6000 to 300 A/mm. at 10,000 A. The instrument was found to be transparent to about 10,000 A; a few lines of the comparison spectrum were observed beyond this, but only weakly, although some of the photographic plates used were believed to be sensitive even further into the infra-red.

For some of the investigations a grating-on-prism spectrograph giving a dispersion of about 25 A/mm. was used. This instrument was only suitable for wavelengths shorter than 8000 A. Some observations were also made on a large Littrow type glass prism spectrograph; this gave a dispersion of 18 A/mm. at 7160 A, but long exposures were required and some difficulty was experienced with temperature control, although one useful spectrogram (plate 7, x) was obtained.

The ordinary iron arc was not found to be a suitable comparison spectrum for the infra-red when working with small dispersion, continua from the hot poles and band structure tending to obscure the line spectrum. For the region 6000–8500 A a neon glow lamp was used for most of the work. This was a rather faint but very simple and reliable source for use as a comparison spectrum. For work further into the infra-red a barium arc (barium nitrate on carbon poles) was selected after various trials. This source gave a relatively small number of very strong lines in the infra-red and was not marred by continuous emission. Wave-lengths measured by Harrison (1939) were used. This barium arc appears to be suitable as a comparison spectrum, when working with small dispersion, for all wave-lengths between 5800 and 10,000 A.

A large variety of photographic plates were used to cover the region 5000–10,000 A. The types used are indicated in the description of the spectrograms reproduced. In some cases plates were sensitized before use by bathing in ammonia (60 sec. in 4 %). Caustic hydroquinone developer was used; this gave high contrast.

THE OXY-HYDROGEN FLAME

It is well known that the oxy-hydrogen flame shows strong OH bands in the near ultra-violet and a weak continuum throughout the visible, the origin of this continuum being the subject of investigations at present being conducted by the author. In addition, it has also been found that there are a number of weak bands at the red end of the visible spectrum; these appear to be identical with the bands observed in the flame of oxygen burning in hydrogen and described by Kitagawa (1936); Kitagawa's measurements, which are probably more accurate than those of the author, are reproduced in table 1.

TABLE 1. VISIBLE EMISSION BANDS OBSERVED BY KITAGAWA IN
THE FLAME OF OXYGEN BURNING IN HYDROGEN

λ	I	λ	I	λ	I	λ	I
5683·3	2	5923·8	4	6220·0	5	6490·4	9
5715·3	1	5948·8	6	6255·1	7	6516·8	10
5806·9	3	5988·8	6	6321·6	7	6574·5	8
5861·6	3	6165·7	7	6377·1	7	6628·6	7
5880·2	6	6181·5	4	6457·5	8	6922·0*	2
5900·2	6	6202·6	7	6468·0	9		

* The author's measurement of the head of this band is 6919·0, and the intensity (using suitable plates) appears greater than 2.

These bands, as obtained by the author in the flame of oxygen burning in hydrogen are shown in plate 6, *b*, and as obtained in the oxy-hydrogen flame in plate 7, *i*. The bands at 6165 and 6457 appear outstanding as heads of groups.

By focusing an end-on image of a powerful oxy-hydrogen flame on to the slit of the large aperture spectrograph it was possible to record the spectrum in the region 7000–10,000 A satisfactorily with exposures of from ½ to 10 min., the exposure times in the region 9000–10,000 A being about the same as that required

A. G. Gaydon

for the barium arc comparison, a striking demonstration of the strength of the infra-red emission from the flame. Spectrograms of the oxy-hydrogen flame are reproduced in plate 7, strips *a–h*. The wave-lengths and intensities of the outstanding features of the bands are collected in table 2.

TABLE 2. BANDS IN THE OXY-HYDROGEN FLAME

λ	I	λ	I	λ	I
6919·0	2 H	8974	3	9440	3
7164·5	6 H	9129	4	9485	3
7299	5	9183	4	9559	4
8097	8 H	9277	10 H	9610	4
8916	7 H	9333	7	9669	7 H

H = outstanding head, degraded to longer λ.

Even with the small dispersion of the large aperture instrument it can be seen that the bands possess a fine structure, and are not all identical in appearance, the 7164 band possessing a much sharper head than the 8097 band. The intensity estimates are based on inspection of spectra on various types of plate, but are not in any sense quantitative as it was necessary to use several types of plate to cover the region and their sensitivity changed rapidly with wave-length; in particular, all the plates used had very low sensitivity at 7000 A, and this made an estimate of the strength of the 6919 A band difficult.

A photograph of the 7164 band was obtained on the large Littrow type spectrograph (plate 7, *x*) and this shows the rotational structure of the band partially resolved. Table 3 gives the wave-lengths, wave numbers (in cm.$^{-1}$ in vacuo) and intensities (visual estimates) of the individual lines of this band. These measurements should be reliable to better than 0·5 A or 1 cm.$^{-1}$.

TABLE 3. THE STRUCTURE OF THE 7164 A BAND

λ	ν	I	λ	ν	I	λ	ν	I
7151·6	13979·1	0	7252·9	13783·8	4	7345·0	13610·9	1
7155·9	970·6	0	7257·6	774·8	1	7349·6	602·4	5
7159·3	964·0	0	7262·2	766·1	2	7354·1	594·1	2
7164·5	953·9	head	7265·2	760·4	2	7358·6	585·8	2
7170·3	942·5	1	7272·3	747·1	0	7363·1	577·6	1
7174·6	936·1	3	7277·3	737·6	3	7365·4	573·3	3 d
7177·5	928·4	3	7288·9	715·7	2	7367·0	570·3	2 d
7181·9	920·1	2	7292·3	709·4	1 d	7373·4	558·6	2 d
7184·6	914·7	2	7295·1	704·1	3 d	7383·8	539·5	4
7187·1	909·9	2	7300·0	694·8	2	7389·2	529·5	3
7194·0	896·7	1 d	7304·3	686·7	4	7401·0	507·9	5
7201·4	882·4	2 d	7310·2	675·7	2	7409·2	493·1	0
7206·7	872·2	3 d	7317·8	661·5	5	7420·0	473·4	2
7211·9	862·1	2 d	7322·8	652·3	2	7424·3	465·5	2
7218·6	849·2	2	7327·9	642·7	3	7439·5	438·0	3
7228·7	829·9	2	7330·3	638·3	2 d	7461·2	399·0	3 d
7235·6	816·7	3 d	7333·7	631·9	3	7480·4	364·6	2 d
7245·2	798·5	2 d	7341·5	617·5	2	7502·1	325·9	1 d

d = diffuse (probably blend of more than one line).

In addition to these bands and the faint continuous background, the only regular features of the spectrum were the lines of sodium and potassium (5890, 5896; 7665, 7699), but the presence of these elements was not likely to be connected with the observed bands. Occasionally dust from the air introduced the calcium oxide bands into the visible spectrum, but a check experiment in which calcium salts were deliberately introduced indicated that the infra-red band spectrum was not associated with calcium. The strong occurrence of the bands in the flame of oxygen burning in hydrogen (i.e. in the absence of nitrogen) appears to limit the emitter to a molecule consisting of oxygen and hydrogen only. It will now be shown that the bands are indeed due to the vibration-rotation spectrum of water.

THE VIBRATION-ROTATION SPECTRUM OF WATER

Water vapour shows strong absorption in the infra-red, and with sufficient thickness, e.g. atmospheric absorption, the bands of this vibration-rotation spectrum can be followed right up to the middle of the visible spectrum. The bands have been studied by a number of investigators, and Mecke and colleagues (Mecke 1933; Baumann & Mecke 1933; Freudenberg & Mecke 1933) have published detailed measurements and analyses of these absorption bands.

Comparison of the flame bands with the absorption bands shows at once that the bands lie in the same regions of the spectrum, but that they are not identical in detail. This is undoubtedly because of the big difference in temperature between the two sources, the absorption measurements being made at atmospheric temperature, while that of the flame exceeds 2000° C. In absorption the strongest lines are those with low rotational energy near the origin of the band, the strong line structure being centred around the origin and not extending to form definite heads. In emission the very high temperature favours lines of much higher rotational energy, and consequently the band extends farther from the origin in both directions and the lines of the multiple R branches are able to reach a turning point and form a head on the short wave-length side. In table 4 the heads of the flame bands are compared with the origins of the absorption bands and with the shortest wave-length lines of appreciable intensity observed by Mecke; it may be noted that for some bands Mecke records very weak lines extending beyond the head of the flame bands, and that, especially for the 7164 band, weak lines in this region are also observed in emission; these lines belong presumably to weaker satellite branches.

The intensities of the emission and absorption bands are qualitatively in close agreement. The very weak absorption band with origin at 7957 A has not been observed in emission; the absorption band at 6324 A is very weak, but the high sensitivity of the photographic plates in this region favours its observation in emission; the other absorption bands observed by Mecke with origins at 5952, 5924 and 5722 are probably present in emission, as indicated by Kitagawa's measurements, but the structure is too complex to assign individual heads to these bands. The moderately strong emission head at 9669 A does not appear to correspond to any observed absorption band; all the absorption bands involve the

ground vibrational level (0, 0, 0), and it is likely that, in emission, transitions will take place to other vibrational levels; the 9669 band falls approximately into the position expected for the heads of the transitions (1, 3, 0) to (0, 1, 0), (1, 3, 1) to (0, 1, 1) and (1, 1, 2) to (0, 0, 0). It is probable that the apparent doubling of many of the band heads is due to the superposition of these transitions to higher vibrational levels, each absorption band being replaced in emission by a sort of sequence.

TABLE 4. COMPARISON OF H_2O BANDS IN EMISSION AND ABSORPTION

λ of heads in emission	λ of origins (Mecke)	shortest λ lines in absorption	upper vibrational quantum numbers		
			v'_σ	v'_π	v'_δ
6165	6324	——	3	1	1
6457	6524	6475·2	1	3	1
6919·0	6994	6937·7	3	1	0
7164·5	7227	7177·1	1	3	0
8097	8227	8126·8	1	2	1
8916	9060	8930·3	3	0	0
9277	9420	9307·3	1	2	0

For the 7164 band, the measurements of the partially resolved rotational fine structure enable a more detailed comparison to be made with the 7227 absorption band. In the region of the origin there is no agreement, as is to be expected on account of the different temperatures of the sources. Near the head of the band it is, however, possible to identify a number of lines, as indicated in table 5. It will be seen that the agreement is well within the accuracy of the author's measurements, and indeed is sufficiently systematic to indicate that the present measurements are about 0·2 A too high. It is also possible to establish a measure of correlation between the lines in emission and absorption at the long wave-length end of the band, and to form most of the strong lines at the long wave end of the emission band into an extension of Mecke's P branch. Also it is possible to extrapolate the lines of the R branches, observed in absorption, to the turning point corresponding to the head of the band; this gives a value of 7162 ± 3 A, which agrees well with the observed head at 7164·5 A.

TABLE 5. COMPARISON OF LINES OF 7164 BAND IN EMISSION AND ABSORPTION

emission	absorption	J_i–J_k
7177·5	7177·375	5_{-3}–4_{-2}
7184·6	7184·536	4_{-2}–3_{-1}; 4_1–3_2
7187·1	7187·016	4_{-1}–3_0
7194·0	7193·776	4_{-3}–3_{-3}
7201·4	7201·205	3_{-2}–2_{-1}
7206·7	7206·433	2_{-2}–1_{-1}

It is thus clear that the new emission bands in the oxy-hydrogen flame form part of the vibration-rotation spectrum of water. It may be noted that the strongest band with head at 9277 A, which with the 9669 band extends to about

Gaydon

Proc. Roy. Soc., A, volume 181, plate 6

Gaydon

Proc. Roy. Soc., A, volume 181, *plate* 7
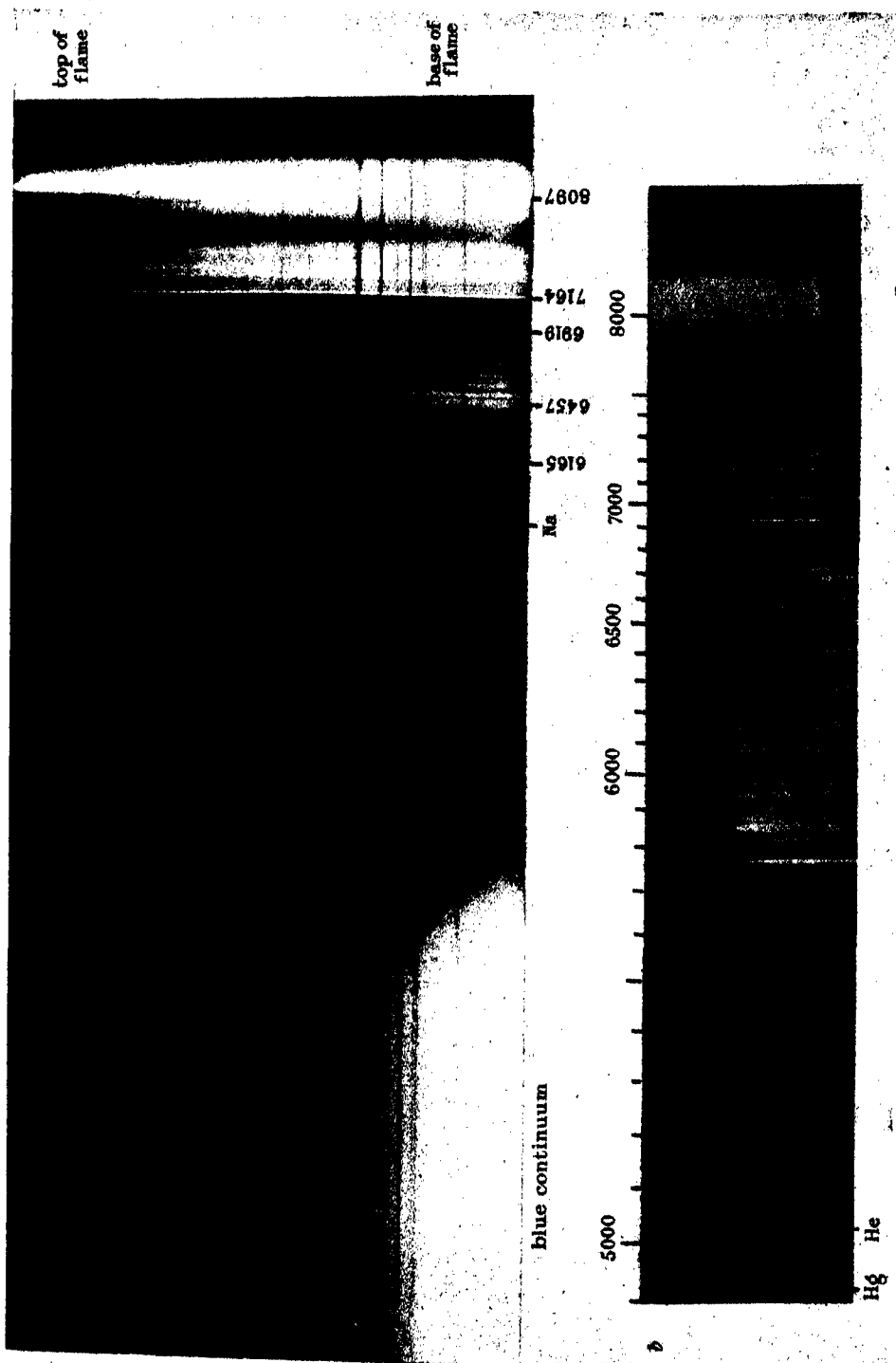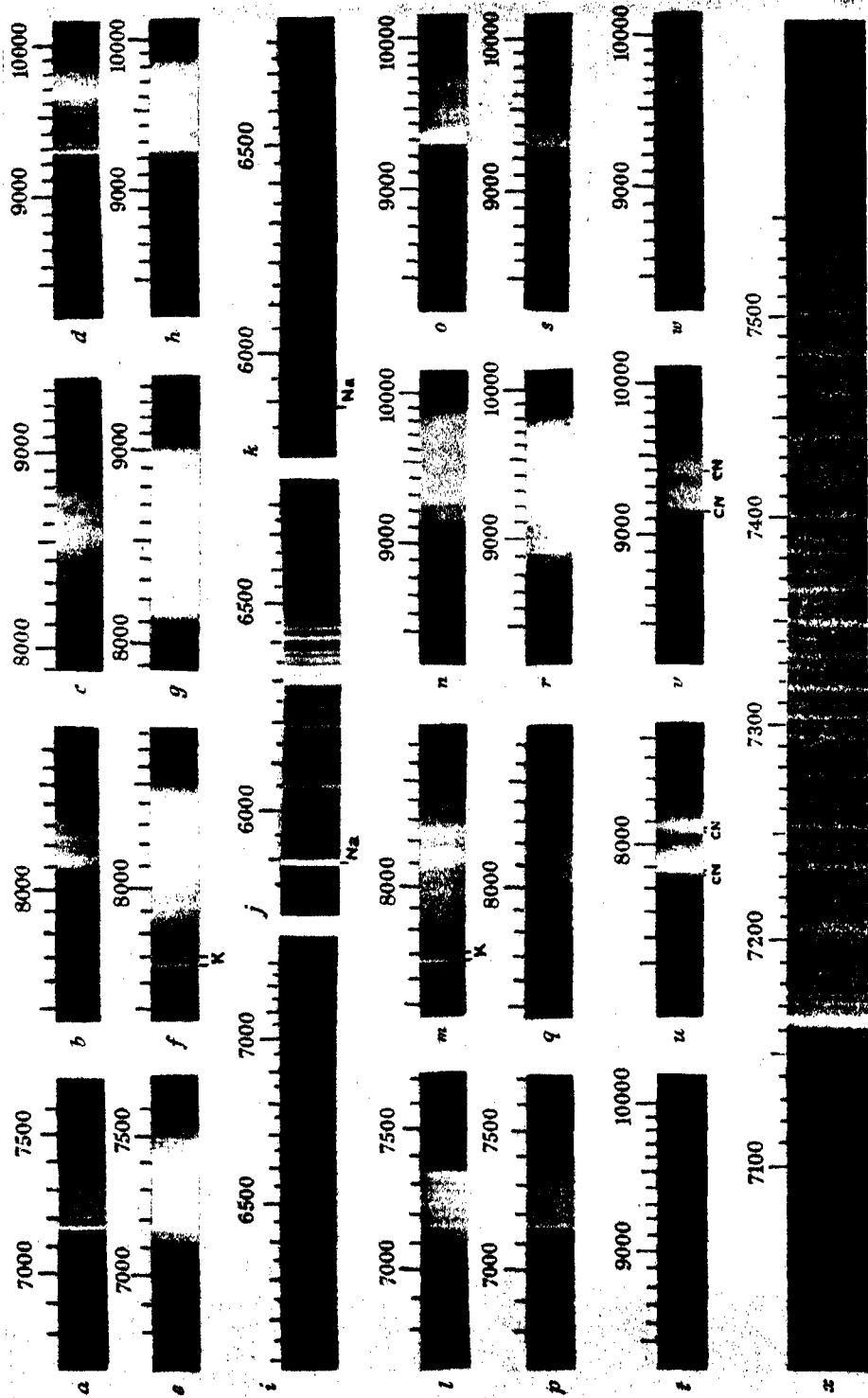
0800 A, corresponds to the band at 0·94 μ recorded by Bailey & Lih (1929) in thermopile measurements on flame spectra, thus linking the bands observed photographically with the known infra-red emission of water vapour in flames.

THE FLAME OF OXYGEN BURNING IN HYDROGEN

The flame of oxygen burning in hydrogen has a bright blue base, and is surrounded by a red mantle. This red mantle is very definite and it is surprising that there does not appear to be any record of its being observed before. The mantle is particularly strong when the supply of hydrogen is reduced, so that it is only in slight excess of that required for the complete combustion of the oxygen; the reason for this is probably that a fast flow of hydrogen, which has a high thermal conductivity, cools the flame. In these experiments the mantle was observed to extend for from 1 to 3 cm. above the main blue flame, the red dying out gradually without any sharp edge. A spectrogram of the flame, taken on a long range spectrum plate, is shown in plate 6, *a*. It will be seen that the blue base of the flame shows a continuum throughout the blue part of the spectrum which dies out about half-way up the flame. The new water emission bands extend, however, throughout the whole length of the flame, decreasing in intensity with height. The red mantle is obviously produced by this emission of the vibration-rotation spectrum of water.

This water-vapour emission appears to be largely thermal in character for flames of oxygen and hydrogen. This is indicated by the effect of excess hydrogen in cooling the flame and so reducing the red mantle, an effect which can less readily be explained if the excitation were chemical. The red glow extends for over a centimetre, and as the rate of flow of gas in most of the experiments was only around 50 cm./sec. it would be necessary to assume a life of over $\frac{1}{50}$ sec. for the vibrationally excited molecules if the emission were attributed to chemical causes. There does not appear to be any evidence from measurements of supersonic dispersion that water vapour possesses any relaxation time comparable with this.

The flame of hydrogen burning in oxygen is also blue, but does not show the red mantle nearly so clearly. Photographs of the spectrum of the flame do reveal, however, that the emission of the new water bands in the photographic infra-red extends slightly above the blue continuum.

THE HYDROGEN—NITROUS-OXIDE FLAME

The visible and ultra-violet spectra of flames of hydrogen burning with nitrous oxide have been described by Dixon & Higgins (1926) and Fowler & Badami (1931). With pre-mixed gases the flame has a non-luminous outer cone, which shows the OH bands, and a brightly luminous yellow cap to the inner cone. The spectrum of this yellow cap shows bands in the ultra-violet due to NO (γ system) and NH, and in the visible a complex many-line structure which was identified by Fowler & Badami with the ammonia α band (Rimmer 1923) which is probably emitted by the radical NH_2. This structure may be seen in plate 7, *j*. It may be noted that,

in the plate, structures at about 6300 and 6330 A are very outstanding. These appear to vary somewhat in intensity from plate to plate, and the explanation of this is not obvious. The structures do not appear to coincide with any lines or bands which might be expected as impurities. It may be that the structures are not of the same origin as the rest of the α band, or that for some reason they are particularly sensitive to changes in the temperature of the flame.

As the proportion of nitrous oxide in the pre-mixed flame is increased the yellow cap to the inner cone becomes less noticeable, and the whole flame assumes a greyish green colour. The spectrum of the flame then appears to be continuous. The colour of the flame closely resembles that of the air (or so-called oxygen) afterglow. As already stated, a study is at present being made of the cause and nature of continuous spectra in flames.

In the infra-red the flame, in the condition when the yellow cap to the inner cone is prominent, shows strongly the new bands attributed to water at 7164, 8097, 8916, 9277 and 9669 A, as may be seen from strips l, m, n and p of the plate. In addition the inner cone shows some other structure, apparently with open rotational spacing, which is strongest in the nearer infra-red, extending to perhaps 8400 A, but is absent, or at any rate weak beyond this. The most obvious feature is an apparent head, degraded to the violet, at 7350 A. A series of lines or maxima of intensity at about $\lambda\lambda$ 7099, 7166, 7207, 7241, 7274, 7304, 7329 close up to form this head at 7350 (see strip l). There is also band structure between 8200 and 8270 A superposed on the new water band (see strip m) and other weak band structure throughout this region. It seems likely that it is due to an extension of the ammonia α band.

With excess of nitrous oxide, when the flame is greyish green in colour, these bands in the infra-red, like the ammonia α band in the visible, are replaced by continuous spectrum. See strips q and r compared with m and n. The new water bands are definitely weaker under this condition, and the bands at 7164 and 8097 were barely observed against the continuous spectrum, but the stronger bands at 8916, 9277 and 9669 A were still easily recognisable despite the continuum, which extended even to these long wave-lengths (see plate 7, r).

The flame of nitrous oxide burning in hydrogen also showed this continuum, and the infra-red bands due to water, but the ammonia α band was not observed. The flame of hydrogen burning in nitrous oxide showed only continuous emission.

THE CARBON-MONOXIDE—OXYGEN FLAME

The visible and ultra-violet spectrum of the carbon monoxide flame was studied by Weston (1925) and later by Kondratjew (1930) and the author (1940). The blue flame shows strong emission consisting of numerous close bands, which are so merged as to appear almost continuous; the emission is strongest in the blue and violet, but extends weakly from the red to far in the ultra-violet (2500 A). The OH bands are very persistent and can only be eliminated, in the case of the flame at atmospheric pressure, by the most careful drying.

With carbon monoxide which has been stored in steel cylinders weak iron lines and FeO bands are observed in the spectrum, the latter being sometimes so strong as to cause the flame to become yellow and resemble a luminous 'carbon-deposition' type of flame. The iron (in the form of the carbonyl) can be removed by passing the gas through a heated tube packed with broken porcelain. This was done with great care in all these experiments as bands due to FeO are also known in the infra-red (Pearse & Gaydon 1941). Comparison plates in which iron was deliberately allowed to be present, so as to show these infra-red bands, were also taken to eliminate any risk of their being spuriously attributed to the flame of the pure gas.

Throughout the photographic infra-red the carbon-monoxide—oxygen flame did not show any distinctive new feature. There was a certain amount of continuous emission in the shorter wave-length part of the region studied. With the small dispersion used it was impossible to say whether this was truly continuous or was an extension of the faintly banded spectrum which is so strong in the blue and violet. In addition the new water bands were present; they were, of course, less strong than in the hydrogen flame, but were considerably stronger than might have been expected from calculations of the amount of moisture present. The bands at 7164 and 8097 A were definitely present although slightly masked by the continuum. The bands at 9277 and 9669 A showed up clearly. In the reproduction (plate 7, *o*) the head at 9277 A is outstanding; the plate used for this spectrogram was Kodak type 144 Q, and the 9669 head does not show up as clearly with this type of plate as on type 144 M; other spectrograms, which are not reproduced, show that the 9277 and 9669 bands had about the same relative intensity as in the hydrogen flame.

In view of the apparently surprising strength of these water bands in the carbon monoxide flame some approximately quantitative estimate of the relative strength of the bands in the carbon monoxide and hydrogen flames seemed desirable. Accordingly a series of exposures with the same slit width were taken for the two flames, the exposure times and approximate gas flows being noted. Exposures of about 30 and $\frac{1}{4}$ min. for the carbon monoxide and hydrogen flames, respectively, were required to lightly record the 9277 A band. The consumption of carbon monoxide was about 0·7 cu. ft./hr., and of hydrogen about 12 cu. ft./hr. The pressure of the gas in both the oxygen and carbon monoxide cylinders during these experiments was about 20 atm. If it is assumed that the gases in the cylinders were saturated with water at room temperature (around 15° C.), then from the vapour pressure of water at that temperature it follows that the gases would contain about 0·09 % by volume of water vapour. With the oxy-hydrogen flame two parts of hydrogen will combine with one of oxygen to give two of water vapour, so that the amount of moisture formed would represent 67 % of the volume of gases burnt. From these figures it follows at once that each molecule of water in the carbon monoxide flame emits about 100 times as strongly as each molecule in the hydrogen flame, a rather remarkable result.

In these experiments the carbon monoxide flame was shielded from air, in order that moisture in the air might not invalidate the results. The numerical value obtained cannot be of high accuracy; first, the measurements of intensity were

somewhat rough; secondly, the sensitization of the plates with ammonia probably varied somewhat; thirdly, the temperature of the two types of flame was very different, the oxy-hydrogen flame being presumably much hotter; fourthly, the flames were of different shape and size. Nevertheless, the result, of the order 100, clearly indicates that each water molecule present in the carbon monoxide flame does radiate much more strongly than corresponding molecules in the hydrogen flame. This agrees with observations by Garner and colleagues (see Gaydon 1942 b) on the spectra of explosion flames in the far infra-red.

Thus it is clear that the radiation of the vibration-rotation spectrum of water from the carbon monoxide flame is not thermal in character. It must result from the chemical process of the combustion. The important effect of water on the combustion of carbon monoxide is well known. At high concentrations of water the combustion probably takes place almost entirely through the maintenance of the water-gas equilibrium. At small concentrations of moisture the catalytic action of the water is very marked and recently (Gaydon 1940, 1941 a, 1941 b, 1942 b) it has been suggested that, in the combustion of carbon monoxide, vibrationally activated molecules of carbon dioxide are formed. These have a comparatively long life on their own, but are quickly restored to equilibrium by water molecules. The release of the energy stored in the vibrationally activated molecules, and the accompanying suppression of the abnormal dissociation of the carbon dioxide, results in more energy being available for the maintenance of the combustion. In the present experiments we have therefore additional proof that water can remove the energy of the carbon dioxide molecules, being itself activated in the process, but quickly losing the energy so acquired either by collision or by radiation, giving the strong emission of the vibration-rotation spectrum of water which is in fact observed from the carbon monoxide flame.

THE CARBON-MONOXIDE---NITROUS-OXIDE FLAME

The flame of carbon monoxide burning in nitrous oxide was also studied by Weston (1925), who found that the spectrum was similar to that of the ordinary carbon monoxide flame, showing the same faintly banded structure in the blue and violet. The OH bands also appear strongly on Weston's published plate, indicating that they are persistent in the flame with nitrous oxide as in the flame with air or oxygen, and therefore probably showing that water is again of importance in the reaction process. The similarity of the flames in nitrous oxide and oxygen is, however, certainly not the whole truth. The carbon-monoxide—oxygen flame is a clear blue, whereas the flame of either carbon monoxide burning in nitrous oxide or the pre-mixed carbon-monoxide—nitrous-oxide flame is of a very different colour which is rather difficult to describe, but resembles that of a London fog more nearly than the sky blue of the flame with oxygen. It is probable that the colour of the flame with nitrous oxide is partly determined by the normal carbon monoxide flame spectrum, and partly by a greyish green continuum similar to that shown by the flame of hydrogen burning with excess of nitrous oxide.

Throughout the photographic infra-red the pre-mixed flame was found to show only some continuous spectrum together with the new water bands (see plate 7, *s*), which were of about the same strength as in the carbon-monoxide—oxygen flame containing a similar amount of moisture; the continuous emission was rather stronger from the flame with nitrous oxide.

The author is unaware of any experiments on intensive drying of carbon-monoxide—nitrous-oxide mixtures to see whether the presence of moisture is as important to the combustion as it is with carbon-monoxide—oxygen mixtures. The persistence of the OH bands in the spectrum of the flame, and now the observation of a similar persistence of the new water bands seems to show that the presence of water is again important. In the combustion with nitrous oxide it is not possible to invoke the maintenance of the normal water-gas equilibrium to explain the action of water in promoting the combustion. The considerations previously put forward to explain the afterburning and catalytic action of water in the combustion with oxygen (Gaydon 1940, 1941*a*, 1941*b*, 1942*b*) will, however, still apply to the combustion with nitrous oxide. In the formation of carbon dioxide from carbon monoxide it will still be true that an electronic rearrangement of the newly formed CO_2 molecule must occur, and this must again lead to the formation of vibrationally activated molecules. These will then be readily deactivated by collision with water molecules, so setting the energy free to maintain the combustion and at the same time giving proof of the process by the emission of the vibration-rotation spectrum of H_2O.

THE OXY-METHANE FLAME

The spectrum of the ordinary oxy-methane flame is well known in the visible and ultra-violet. The inner cone shows strong bands due to C_2, CH and OH, with a weaker system of bands now known as the hydrocarbon flame bands, which are probably due to HCO (Vaidya 1934, 1941; Gaydon 1942*a*). The outer cone shows the OH bands strongly, and some of the faintly banded structure characteristic of the carbon monoxide flame.

In the photographic infra-red both the inner and outer cones of the flame were found to show strongly the new bands attributed to the vibration-rotation spectrum of water. The 9277 and 9669 A bands may be seen in strip *t* of plate 7. All the bands appear to be identical in appearance with those observed in the oxy-hydrogen flame, and no difference in relative intensity of the bands was obvious, although it must be admitted that such a difference would be difficult to detect because of the need to observe most of the bands separately on different types of photographic plate.

It may be noted here that Bailey & Lih (1929) recorded observing the water band at $0.95\,\mu$ in a hydrogen flame, and at $0.94\,\mu$ with coal-gas in a Bunsen or Meker burner; they did not record the band in a methane flame. Since the band has a sharp head at 9277 A and the maximum of intensity is in this head, it is rather surprising that they record a variation in the wave-length of the band. For the bands further in the infra-red, which correspond to smaller changes of vibrational

quantum number, and therefore to smaller changes of the moment of inertia, it is possible that the heads are less well developed, and that the wave-length of the intensity maximum of the band may vary with the flame temperature. There does not, however, appear to be any evidence of this effect for the bands in the photographic infra-red.

THE METHANE—NITROUS-OXIDE FLAME

The spectrum of the methane—nitrous-oxide flame does not appear to have been described in detail. Dixon & Higgins (1926) have described the appearance of this and other hydrocarbon flames in nitrous oxide, and Barratt (1920), in discussing the origin of the cyanogen bands, noted that a coal-gas—nitrous-oxide flame showed the violet bands of CN and the NH band as well as the usual OH, CH, and C_2 Swan bands. In the present investigations of the visible spectrum of the methane—nitrous-oxide flame it has been found that the inner cone shows the CN violet bands strongly, the Swan bands of C_2, the CH bands, and in addition the ammonia α band. This α band can clearly be seen in strip k of plate 7; comparison of this with strip j, which shows the hydrogen—nitrous-oxide flame, reveals a difference of intensity of the structures at 6300 and 6330 A which has already been commented upon; however, other plates taken of the flames with the grating-on-prism spectrograph do not reveal this marked difference in intensity.

In the infra-red the spectrum of the outer cone is identical with that of the hydrogen flame, showing only the new water bands (see plate 7, w). The inner cone also shows the water bands, but in addition strong bands belonging to the red system of CN are present. These bands have several heads; the R_2 heads of the strongest bands (from Asundi & Ryde 1929) are at 7874 (1, 0), 8067 (2, 1), 9140·5 (0, 0) and 9393 (1, 1). They can be seen in strips u and v of the plate. The relative weakness of the red cyanogen bands, compared with those of the violet system, in flames has recently been commented upon (Gaydon 1942b); it now appears that this is probably more apparent than real, the strong bands of the red system lying in the region of the infra-red which has not previously been studied.

It is a pleasure to express my thanks to Professor A. C. Egerton for his interest in the work and for many helpful discussions.

REFERENCES

Asundi, R. K. & Ryde, J. W. 1929 Nature, Lond., 124, 57.
Bailey, C. R. & Lih, K. H. 1929 Trans. Faraday Soc. 25, 29.
Barratt, S. 1920 Proc. Roy. Soc. A, 98, 40.
Baumann, W. & Mecke, R. 1933 Z. Phys. 81, 445.
Dixon, H. B. & Higgins, W. F. 1926 Mem. Manchr. Lit. Phil. Soc. 71, 15.
Fowler, A. & Badami, J. S. 1931 Proc. Roy. Soc. A, 133, 325.
Freudenberg, K. & Mecke, R. 1933 Z. Phys. 81, 465.
Gaydon, A. G. 1940 Proc. Roy. Soc. A, 176, 505.
Gaydon, A. G. 1941a Proc. Roy. Soc. A, 178, 61.
Gaydon, A. G. 1941b Nature, Lond., 148, 226.
Gaydon, A. G. 1942a Proc. Roy. Soc. A, 179, 439.
Gaydon, A. G. 1942b Spectroscopy and combustion theory. London: Chapman and Hall.

Harrison, G. R. 1939 *M.I.T. wave-length tables.* London: Chapman and Hall.

Kitagawa, T. 1936 *Proc. Imp. Acad. Japan*, 12, 281.

Kondratjew, V. 1930 *Z. Phys.* 63, 322.

Mecke, R. 1933 *Z. Phys.* 81, 313.

Pearse, R. W. B. & Gaydon, A. G. 1941 *The identification of molecular spectra.* London: Chapman and Hall.

Rimmer, W. B. 1923 *Proc. Roy. Soc.* A, 103, 696.

Vaidya, W. M. 1934 *Proc. Roy. Soc.* A, 147, 513.

Vaidya, W. M. 1941 *Proc. Roy. Soc.* A, 178, 356.

Weston, F. R. 1925 *Proc. Roy. Soc.* A, 109, 177, 523.

Description of Plates

All the spectrograms except 7 (*x*) were taken on the large aperture glass prism spectrograph. The type of flame, the exposure times, and the plate types used are given in the following list.

Plate 6

(*a*) Oxygen burning in hydrogen showing both base and top of flame. 28 min. Ilford long-range spectrum.

(*b*) Top of flame of oxygen burning in hydrogen; greater enlargement. 40 min. Neon lamp comparison, 1 min. Ilford long-range spectrum.

Plate 7

(*a*) Oxy-hydrogen. 1 min. Kodak I.U.

(*b*) Oxy-hydrogen. 2 min. Ilford infra-red.

(*c*) Oxy-hydrogen. 15 min. Kodak 144 P.

(*d*) Oxy-hydrogen. 2 min. Kodak 144 M.

(*e*) Oxy-hydrogen. 10 min. Kodak I.U.

(*f*) Oxy-hydrogen. 4 min. Ilford infra-red.

(*g*) Oxy-hydrogen. 30 min. Kodak 144 P.

(*h*) Oxy-hydrogen. 4 min. Kodak 144 Q.

(*i*) Oxy-hydrogen. 2 min. Ilford long-range spectrum.

(*j*) Hydrogen—nitrous-oxide, inner cone. 8 min. Ilford long-range spectrum.

(*k*) Methane—nitrous-oxide, inner cone. 2 min. Ilford long-range spectrum.

(*l*) Hydrogen—nitrous-oxide, inner cone. 6 min. Kodak I.U.

(*m*) Hydrogen—nitrous-oxide, inner cone. 15 min. Ilford infra-red.

(*n*) Hydrogen—nitrous-oxide, inner cone with moderate flow of N_2O. 15 min. Kodak 144 Q.

(*o*) Carbon-monoxide—oxygen (moist). 30 min. Kodak 144 Q.

(*p*) Hydrogen—nitrous-oxide, outer cone. 6 min. Kodak I.U.

(*q*) Hydrogen—nitrous-oxide, inner cone with excess N_2O. 6 min. Kodak I.U.

(*r*) Hydrogen—nitrous-oxide, inner cone with excess N_2O. 3 min. Kodak 144 Q.

(*s*) Carbon-monoxide—nitrous-oxide (moist). 30 min. Kodak 144 Q.

(*t*) Oxy-methane, inner cone. 30 min. Kodak 144 M.

(*u*) Methane—nitrous-oxide, inner cone. 6 min. Ilford long-range spectrum.

(*v*) Methane—nitrous-oxide, inner cone. 20 min. Kodak 144 M.

(*w*) Methane—nitrous-oxide, outer cone. 20 min. Kodak 144 M.

(*x*) Oxy-hydrogen. 14 hr. Kodak I.U. (sensitized in NH_3) on large Littrow type glass spectrograph.

Amino-acid analysis and the structure of proteins

By A. C. CHIBNALL, F.R.S.

The recent speculations of Bermann and Niemann (B.-N.) on protein structure are reviewed in the light of new analytical data for certain proteins. The molecule of edestin would appear to be a system of six peptide chains of like composition and mol. wt. 50,000, the constituent residues of which conform to the B.-N. stoichoimetric rule in that the number of residues of each of the thirteen residue species for which data were obtained was expressible in terms of 2 and 3. The molecule of lactoglobulin is a system of eight or nine peptide chains, not all of which can be of like composition, while that of egg-albumin is a similar system of four chains, in agreement with the recent views of Astbury. The analytical data show that the molecules of the two latter proteins contradict the B.-N. rule, but it is possible that the component peptide chains may conform to it. Insulin (mol. wt. taken as 35,500) appears to be a system of eighteen peptide chains, in agreement with Bernal's deductions from crystallographic data. The conclusion that the molecules of these and other proteins as systems of peptide chains is based in part on titration data and in part on the estimation of free amino-N; new suggestions are put forward as to the way in which the component peptide chains linked together.

Anniversary Meeting of the Royal Society
and Newton Tercentenary Celebrations,
30 November 1942

Before proceeding to the presentation of the Medals awarded for this year, it is fitting that, in accordance with custom, we should briefly recall the lives and the achievements of those whom death has removed from our Fellowship and our Foreign Membership since the last Anniversary Meeting of the Society.

His Royal Highness the DUKE OF CONNAUGHT, who died on 16 January at the advanced age of 91 years, was elected to our Fellowship under Statute 11 as long ago as 1906, so that he was by many years the senior of the members of the Royal Family who have accepted election to our Fellowship. His Royal Highness'had other contacts with the advancement and the applications of scientific knowledge as President of the Royal Society of Arts, and of the Royal Colonial Institute. The latter appointment had a particular fitness in recognition of his many and great services to the British Commonwealth of Nations, as Governor-General of the Dominion of Canada and on other special missions.

EMILE PICARD (1856–1941), perpetual secretary of the Paris Academy of Sciences since 1917, had been a Foreign Member since 1909. He was one of the most famous of modern French mathematicians. He did work of fundamental importance in the theory of differential equations, the theory of algebraic functions of several variables, and the theory of surfaces and integrals associated with them; and 'Picard's theorem' is one of the classical theorems of the theory of analytic functions.

TULLIO LEVI-CIVITA (1873–1941) was elected a Foreign Member of the Society in 1930, and represented the brilliant school of mathematics in Rome. Gifted with a remarkable command of the resources of mathematical analysis and with a keen geometrical intuition, he applied his great technical skill to nearly all branches of mathematics, but it was to problems arising in applied mathematics that he devoted most of his efforts. Electrostatics, analytical dynamics, celestial mechanics, hydrodynamics, relativity, and quantum mechanics, as well as problems in engineering, all claimed his attention in turn. Mathematicians, however, associate Levi-Civita first of all with the Absolute Differential Calculus, which he helped to found with his teacher Ricci. It was in 1917 that Levi-Civita advanced the subject greatly by introducing the idea of parallel displacement. This discovery, following on the announcement of Einstein's general theory of relativity, in which the absolute differential calculus had proved a most powerful tool, provided a means of making important advances in relativity theory, culminating in the unified

theory of Weyl. Levi-Civita's numerous text-books are models of lucidity, and two of his treatises, *The Absolute Differential Calculus* and *Rational Mechanics*, are among the leading works on the subjects with which they deal.

In the EARL OF BERKELEY (1865–1942) the Society has lost a Fellow of versatility and distinction, who used his private fortune to build and equip a laboratory in which he made an outstanding contribution to the measurement of osmotic pressure. He had shown promise of a fine career in the Navy, but he left the Service when he was 22 to devote himself to research. He soon became a brilliant experimenter, and the absence of any formal training was in some ways an advantage to a man with his instinctive grasp of scientific method, as his approach to every problem was along original lines. In the great days of the Foxcombe laboratory he had a well-balanced team of workers, including Ernald Hartley, Charles Burton and M. P. Appleby, and he proved himself a fine leader with initiative and imagination. His work on osmotic pressures gave ample scope to his engineering and mathematical ability, and the agreement of the directly observed values with those calculated from vapour-pressure measurements established once for all the validity of the indirect method.

Sir WILLIAM HENRY BRAGG (1862–1942), who died on 12 March in his eightieth year, had been our President for a period so recent as 1935–40. Even more recently, during my own absence in America, he had come, from retirement, to occupy this chair as Vice-President for the two months from 20 December to 20 February last, and thus to give again to the Society's service the full measure of his wisdom and experience, up to a few weeks before a brief illness brought the end. He occupied a very special place in the scientific life, not only of the Royal Society, but of the whole nation, and, indeed, of the world at large. Without any compromise of the highest standards in his experimental and theoretical researches, he had made himself supreme as an interpreter of the results of science to those least equipped with special knowledge; so that he seemed equally happy and assured when discussing the complexities of crystal structure with fellow experts, and when presenting and enlivening the elements of science before a fascinated audience of children.

From Trinity College, Cambridge, and a high place among the Wranglers, he went in 1887 as a young man to Adelaide, to occupy a new chair of mathematics and physics. There followed a period of many years devoted to teaching and organizing, and his career as an experimental investigator, in which he attained the highest rank, was remarkable, if not unique, in the fact that it did not begin till he was 42 years of age. It was not, indeed, until he had returned to England and to the chair of physics at Leeds that, at the age of 46 years, he began, in collaboration with his son William Lawrence Bragg, the researches on crystal architecture as revealed by the reflexion of X-rays from crystals, which established his fame. They also provided the opening to the long series of masterly investiga-

tions, in which a succession of brilliant collaborators worked under his guidance and inspiration for some years at University College, London, and then, for the rest of his life, at the Royal Institution and its Davy-Faraday Research Laboratories. Here Bragg found the ideal setting and opportunity for the exercise of his distinguished gifts for research and for popular exposition, and of the hospitality and kindliness which endeared him to so wide a circle.

JEAN PERRIN (1870–1942), who died in New York on 17 April, had been Professor of Physical Chemistry at the University of Paris for about forty years. He was one of the best-known physicists of his day and made valuable contributions to the knowledge of atomic physics, both by his experimental work and by his books on modern physics. In 1926 he was awarded the Nobel prize for the remarkable work he had done on the Brownian movement of small particles suspended in a liquid. By a series of newly devised methods of measurement he succeeded in making a definite estimate of the mean energy of agitation of the particles, and thence deduced the number of molecules per cubic centimetre of a gas at normal temperature and pressure. He had many interesting views on physical phenomena which are described in an attractive manner in his books, *Les atomes* and *Les éléments de la physique*. In a later publication, *Grains de matière et de lumière*, he discusses the modern theory of radiation and ionization.

Perrin also took a great interest in scientific and industrial research, and in his capacity of scientific adviser to the French government he was largely responsible for the organization of research in applied physics and the establishment of the Palais de la Découverte.

GEORGE GERALD STONEY (1863–1942), a life-long friend and colleague of the late Sir Charles A. Parsons, was world famous as one of the best-known pioneers of the steam turbine and high-speed dynamo electric machine. Educated privately and later at Trinity College, Dublin, he was one of the original staff of the steam-turbine and electrical works founded by Parsons at Newcastle-on-Tyne, and was the sole survivor of the original crew of the experimental steam yacht *Turbinia*.

From the foundation of Heaton Works in 1889, Stoney ably assisted Sir Charles in the long struggle that lay ahead to achieve recognition of the merit of the compound steam turbine, till it became the essential instrument for providing power on the largest scale. Stoney had the privilege of being an active witness of this epic of engineering history, and although he sought no publicity or public recognition of his work he became almost as well known in the world of power generation as Sir Charles himself. Content simply to support his chief to the best of his ability, Stoney contributed many papers to engineering societies and journals. He became a Fellow of the Society in 1911.

Sir JOSEPH LARMOR (1857–1942), Copley Medallist and for eleven years Secretary of the Society, was Lucasian Professor of Mathematics at Cambridge from 1903

to 1932. His distinguished contributions cover many branches of mathematical physics and geophysics. He will be especially remembered for his fundamental researches on the relations of matter, electricity and ether. These, together with the contemporary work of Lorentz, are the bridge which unites the revolutionary theories of the present century to the older physics. To Larmor modern atomic theory owes the 'Larmor precession' and the formula for the radiation of an accelerated charge, which are the basis of its most fertile developments. As an Irish mathematician he had strong attachment to the school of Hamilton, MacCullagh and FitzGerald, and he worthily carried on its great tradition.

The Society lost its senior Fellow by the death on 2 June of ANDREW RUSSELL FORSYTH (1858–1942), who had been elected at the age of twenty-seven in 1886. After graduating as Senior Wrangler in 1881, he produced in rapid succession a series of important memoirs on Theta-functions, Abel's Theorem, and Invariant-theory. In 1893 he was elected to the Sadleirian chair at Cambridge in succession to Cayley; in this position he rendered services of the greatest value to British mathematics by making known, in lectures and advanced treatises, the fruits of continental research. He was an exceptionally able administrator, and took a prominent part as a reformer in many academic movements and controversies.

Forsyth left Cambridge in 1910, and from 1913 to 1923 was Chief Professor of Mathematics in the Imperial College of Science and Technology. In 1897 he was awarded the Royal Medal: and in the course of his long life he received very many honours from Universities and foreign Academies. His devotion to the Royal Society, and to the promotion of science in fields far distant from those in which he was active himself, is shown in the bequest of his estate to the Society for the support of medical researches.

ALFRED DANIEL HALL (1864–1942) probably did more for the introduction of science into modern agriculture than any other man in our time. He possessed a many-sided ability and a remarkably wide range of interests, including Japanese prints, tulips, music, poetry and French literature; and he was a gardener of distinction. After leaving Oxford he spent a few years as schoolmaster and University Extension Lecturer. He then started the Agricultural College at Wye, Kent, where he worked out courses of instruction far ahead of any then existing. Later he took charge of Rothamsted and made important investigations on soils and crop production. Facilities, however, were very inadequate, and, realizing that more must be provided, Hall transferred his activity and his centre of interest to the Development Commission, where he played an important part in inaugurating the system of agricultural education, advice and research that has contributed so greatly to the advancement of the science and the industry.

WILLIAM HENRY YOUNG (1863–1942), Sylvester Medallist of the Society in 1928, was one of the most profound and original among British mathematicians of the

last fifty years. He was one of the leaders in the development of the modern Cambridge school of analysis, and there are few branches of the theory of functions on which he has not left his mark. The theory of sets of points, the foundations of the differential and integral calculus, the theory of Fourier series and other orthogonal developments, are full of striking theorems discovered by Young.

WILLIAM MATTHEW FLINDERS PETRIE (1853–1942) was one of the greatest figures the world has seen in the comparatively young science of Egyptology. In his own branch of that science he was pre-eminent. By his excavations in Egypt over a period of forty years he contributed more than any other single scholar in the last half-century to our knowledge of the history of the land in ancient times, and to the use of that history as the chronological yard-stick for the ancient world in general, prior to the age of classical Greece. Above all he was the founder of the scientific method of modern archaeology.

Owing to delicate health in childhood Petrie was educated privately, and largely by finding in his own way the line of his own interests—geology, surveying, chemistry and history. He first went to Egypt to survey the Great Pyramid in 1880. He returned in 1883 to start excavating for the Egypt Exploration Fund, but by 1886 he was working there on his own account, and remained for the rest of his life virtually his own master. In 1892 he was appointed to the Edwards Professorship of Egyptology at University College, London, the first chair in the subject to be founded in this country, from which he retired in 1933. He was elected F.R.S. in 1902, and knighted in 1923.

RICHARD WILLSTÄTTER (1872–1942) was an illustrious organic chemist, distinguished alike for the success that attended his attack on the most difficult problems and for the improvements that he made incidentally in the technique of investigation. His analytic and synthetic skill were shown in the determination of the constitutions of atropine and cocaine and the disclosure of their relation to *cyclo*heptane. Work on aniline black and the orthoquinones opened up new vistas. Having isolated the chlorophylls, he made the surprising discovery of their magnesium content and degraded them, step by step, to porphyrins analogous to those obtainable from the blood pigments. He made important contributions to our knowledge of photosynthesis. Typical of his genius was the isolation of the anthocyanin pigments of fruits and blossoms, their recognition as glycosides and the elucidation of the structure of the latter. In later years he devoted the whole of his energies to biochemical work, for example, to the concentration and separation of the enzymes. Willstätter's premature retirement was the result of a refusal to compromise with forces and tendencies alien to his whole mentality.

EDWARD FAWCETT (1867–1942), who died suddenly at Bristol on 22 September, had a distinguished career as an anatomist. He received his medical education at the University of Edinburgh, graduating M.B., C.M. in 1889 and M.D. in 1906. After a period as Demonstrator of Anatomy at the Yorkshire College, Leeds, he

was appointed in 1893 to the Professorship of Anatomy in University College, Bristol. In spite of a heavy burden of administrative work, Fawcett found time to carry out much original research and came to be recognized as a foremost authority on the morphology and development of the mammalian skeleton, including that of man. In particular, he published in the *Journal of Anatomy* a series of papers, based on accurate wax reconstructions, in which he greatly extended our knowledge of the development of the chondro- and osteo-cranium in various species of mammals. A man of very varied interests, Fawcett in his later years became keenly interested in church architecture. He was elected a Fellow of the Society in 1923.

The death of GEORGE GERALD HENDERSON (1862–1942) has taken from us the doyen of Scottish chemists and closed a career of remarkable usefulness and influence. His thirty-five years of professorial life were spent first in the chair of chemistry in the Royal Technical College, Glasgow, and afterwards as Regius Professor in the University of that city. When he retired he left behind him two vigorous schools of chemistry, which he had consolidated through his zeal for research and his capacity to quicken enthusiasm in young men. Henderson was a brilliant lecturer, an able administrator, and a versatile scientific inquirer. Between his first paper on Dolomite and his last on the Carophyllene Series are found more than sixty valuable publications, most of which deal with the chemistry of the terpenes; but almost any type of problem—organic or inorganic, academic or technical—appealed to his seeking mind and yielded to his skill. He became a Fellow of the Society in 1916 and occupied with dignity and distinction the Presidency of the Chemical Society and of the Society of Chemical Industry.

By the death of JOHN NORMAN COLLIE (1859–1942) on 1 November chemistry has lost one of its most outstanding personalities. To his numerous friends and old students the memory of him will be one of a great organic chemist endowed with a singular vision and understanding.
    Collie was characterized by a remarkable versatility, for he excelled as a mountaineer and an art critic, as well as an authority on early printing and incunabula, Chinese porcelain and lacquers, and Japanese netsukes and sword-guards. He was a true aesthete, for beauty made a strong appeal to him, whether in natural form or in the handiwork of man.
    To those who knew him, Collie was gifted with an understanding of human nature which made him an ideal teacher. Although he was somewhat hesitating in his style, his lectures were singularly attractive, because there always appeared to be something of absorbing interest beyond the horizon of his description.

Award of Medals, 1942··

The COPLEY MEDAL is awarded to Sir ROBERT ROBINSON. He is recognized in all countries as one of the world's leaders in organic chemistry, and is one of the greatest and most versatile of investigators in that department of science. His researches, with a long and notable succession of pupils and collaborators, have covered a remarkably wide range of problems in this field, and his, approach to these has been distinguished by brilliance in conception and a genius for the selection of methods leading to the desired solutions.

Robinson's investigations have been particularly concerned with the chemistry of natural substances, products of the life processes of plants and animals. His work has thus been a potent factor in the tendency of organic chemistry to return, in recent years, to an objective nearer to that of its origin, and to make contacts of growing intimacy and value with biochemistry, a more recent development in response to the stimulus of functional biology.

This occasion does not permit any attempt at a complete or detailed survey of all the different fields which Robinson's work has illuminated and opened to further exploration. Special mention must be made, however, of his long series of fundamental investigations on the constitution and relationships of the plant alkaloids. His theory of the biogenesis of plant products seems rather to have inspired than to have resulted from his own early and elegant synthesis of tropinone; and it has revealed an unforeseen and coherent relationship between the constitutions of different groups of alkaloids, and given a great stimulus to work on their synthesis. The work published from Robinson's laboratory has been fundamental to understanding of the isoquinoline and the indole series of alkaloids, of morphine and its allies, and of the structural formulae of strychnine and brucine, which formed the subject of his Bakerian Lecture.

Of at least equal scientific rank is the work which Robinson carried out and inspired over many years on the anthocyanin and, more recently, on the anthoxanthin pigments of plants, culminating in the synthesis of the actual colouring matters of flowers, and forming as a whole one of the most brilliant achievements in the whole range of modern organic chemistry.

Robinson's mastery of synthetic resources, and his penetrating instinct for clues to organic constitution, have been further demonstrated in a more recent approach to the synthesis of the steroids, in the production of series of compounds of interest for chemotherapy, and in notable studies of individual natural substances of a range of other types. He is, moreover, a philosopher as well as a master of experimental possibilities; and his theory of organic reactions, in the modern, electronic terms of valency bonds, has had a great influence on the development of fundamental conceptions in organic chemistry.

The Copley Medal is the highest recognition of scientific achievement in the Royal Society's gift, with no limitations of subject or nation; and the Society may well find cause for satisfaction in the knowledge that the award of its premier

Medal for achievement in organic chemistry, after an interval of many years, finds among its own Fellows a recipient of such unquestioned pre-eminence as Sir Robert Robinson.

The RUMFORD MEDAL is awarded to Dr G. M. B. DOBSON. The Rumford Medal awarded by the Royal Society was established for the recognition of important discoveries made in Europe, especially on heat or on light. These conditions appear to be met with a special fitness in the award of the medal this year to Dr G. M. B. Dobson, for his meteorological researches and discoveries. For Dobson's work has, in recent years, greatly extended knowledge of the linkage between the behaviour of ozone and cyclonic disturbances, in that complicated heat engine which is the earth's atmosphere; while light may be said to have provided the basis of measurement for Dobson's spectrographic studies of the distribution of ozone, in time and in height above the earth's surface. Light had also a major concern in earlier researches on meteors in which Dobson collaborated with Professor Lindemann (now Lord Cherwell); in these, the study of the heights between which meteors become luminescent enabled them to draw conclusions as to the density, and from these as to the temperature of the atmosphere at great heights from the earth's surface. But it is especially on Dr Dobson's own studies of the ozone of the atmosphere, continued over many years, and producing results of outstanding importance for meteorology, that the award is based.

A ROYAL MEDAL is awarded to Professor W. N. HAWORTH, a brilliant leader in organic chemistry. Haworth's great claim to distinction arises from the revolutionary change which has been produced by his own investigations, and by those of his immediate pupils, in the whole aspect of an important section of organic chemistry, dealing with the structure and the relationships of the carbohydrates. The ring structure of the simple sugars, first proposed by Tollens and supported by the work and the authority of Emil Fischer, had been generally regarded by chemists as firmly established. Detecting insecurities in the arguments which led to this formulation, Haworth developed the methylation technique, first used by Purdie and Irvine, and applied it systematically to the monosaccharides. He was thus able to show, by unequivocal methods of organic chemistry, that the accepted ring structure of these sugars was incorrect, and that, in their normal and reactive forms, they were derivatives of pyran and of furan respectively.

Later Haworth further developed his methods in application to sugars and carbohydrates of increasing complexity. By his work, and that of others who have followed his lead, the detailed structures of many disaccharides and of some trisaccharides have been established. Progress has further been made in his attack on the structural complexities of the polysaccharides, and a simple chemical method has been evolved for determining their minimal molecular weights. Professor Haworth's work, in the field which he has thus made his own, has received the high international recognition of a Nobel Prize, and will assuredly take rank as a major achievement of permanent significance in chemical history.

A ROYAL MEDAL is awarded to Dr W. W. C. TOPLEY, who is one of the most distinguished of the British bacteriologists of recent years. The most important of Topley's contributions to bacteriology and experimental medicine has been the experimental study of epidemics, which he initiated and of which the methods have been largely his own creation. Much had been learned by the statistical analysis of observational data, dealing with the origin, spread and development of natural epidemics, under conditions largely out of control. Topley conceived the idea of applying such methods to the investigation of epidemics started artificially in populations of healthy mice, kept in the laboratory under conditions which could be exactly controlled and deliberately varied. Thus the factors conducive to the rise, culmination and decline of an epidemic, to its revival or its final subsidence, could be experimentally determined. In a long series of such studies, the important results of which were reviewed in his Croonian Lecture for 1941 on *The Biology of Epidemics*, Topley had the statistical co-operation of Professor Major Greenwood, in both planning and interpretation.

On this firm basis of knowledge concerning the incidence and mortality of a naturally transmitted infection in untreated stock, the efficiency of prophylactic measures could be put to a controlled test. Under Topley's guidance and inspiration, accordingly, substantial progress had been made by his chemical colleagues towards the isolation from various species of pathogenic bacteria of highly purified and stable antigens, and the practical trial of some of these was interrupted by the outbreak of war.

Dr Topley's researches have had a great and lasting influence on the study of bacteriology, immunology and epidemiology in relation to human medicine. His recent change of the focus of his interests may be expected to give an important stimulus to advance in many cognate fields of agricultural research.

The DAVY MEDAL is awarded to Professor C. N. HINSHELWOOD for his work on the kinetics of chemical change, characterized by its pioneering quality and by the varied new lines of research which it has opened up. An experimental investigator of great skill and achievement, Hinshelwood has also enlarged the theory of the subject by able mathematical analyses and descriptions based on the concepts of collisions and of activation energy.

Hinshelwood took a leading part in the early study of homogeneous gaseous reactions. As the result of the examination of a number of bimolecular examples he was able to show, with reason, that these are confined to molecules containing few atoms, and that the actual rate is given by the product of the total collision rate and the probability of a molecule possessing the experimental energy of activation. Unimolecular reactions were found to occur with polyatomic molecules and to show more complex features. In association with Professor Lindemann (now Lord Cherwell), Hinshelwood put forward the mechanism, now accepted, whereby a reaction fundamentally dependent on collisions may nevertheless have unimolecular

kinetics. This theory he was able to verify by showing that the rates of such reactions diminish at low pressures and that the kinetics then become bimolecular.

In the field of chain reactions Hinshelwood opened up new lines of advance by studying the thermal reaction between hydrogen and oxygen. Thus he discovered, and offered clear explanations for, the curious phenomenon of 'explosion limits', confining explosive reaction, at any fixed temperature, sharply to a particular pressure region. Elaborate studies of the effects of nitrogen peroxide and other foreign gases on the hydrogen-oxygen reaction brought to light the very great kinetic complexities of an apparently simple type of chemical change. In this work Hinshelwood drew attention to the influence of the container surfaces on chain reactions, and also clarified the confusion of evidence concerning the effects on reaction rates of the intensive drying of gases. He discovered the inhibition of certain gaseous reactions by nitric oxide and interpreted the effect as due to the removal of radicals from, and the suppression of, 'chains'.

Hinshelwood has also carried out a large number of experiments on heterogeneous reactions and shown that their differences in kinetic behaviour can be explained by the application of the concepts of Langmuir.

Throughout all these researches, carried out with the utmost economy and directness, though with full experimental precautions, and interpreted in the most lucid manner, Hinshelwood has never lost sight of the essential complexity of chemical reaction mechanisms. He has always been ready to modify his views in accordance with new experimental evidence, and to make the fullest use of the more recent developments of wave-mechanics and of statistical mechanics. Summarized by their author in two well-known treatises, Professor Hinshelwood's distinguished researches furnish abundant ground for the award to him of the Davy Medal.

The DARWIN MEDAL is awarded to Professor D. M. S. WATSON, pre-eminent among palaeontologists for his contributions to knowledge of the course of vertebrate evolution. His researches have been concerned mostly with the origin of the land vertebrates, with the fishes most nearly related to them, and with the main line of evolution leading to the mammals.

It will not be possible on this occasion to survey Watson's work in all its aspects, and mention must be restricted to some of the major lines of advance which it has opened. His Croonian Lecture, in 1925, summarized the conclusions which he had reached by that date as to the evolution of the Amphibia, demonstrating for the first time the relationship of the Stegocephalia to the Osteolepid fishes.

In addition to tracing the descent of land vertebrates thus from Amphibia back to fishes, Watson followed the line of the evolution of the mammals, through early, primitive reptiles, the Cotylosaurs, to the mammal-like reptiles, in a large series of valuable papers. He related this work on the reptiles to that on the amphibians in a paper on the evolution of the shoulder girdle of the Tetrapoda.

In this work on the fossil vertebrates, in its relation to the course of evolution,

Watson has not confined his attention to morphological details, but, with an enterprise remarkable in a palaeontologist though characteristic of his outlook, has considered where possible the functional significance of the structures preserved in the rocks; as in the paper in which he considers the mode of action of the shoulder girdle and deduces the nature of the musculature of a group of marine fossil reptiles.

Pursuing his study of mammalian origins, Watson was led to study the most primitive of living mammals, the oviparous Monotremes, and to discover that characters in which their skulls differ from those of other mammals can be regarded as extreme developments of features observed in the skulls of certain fossil, mammal-like reptiles.

Watson's work has continued in full vigour into recent years, and has brought two further contributions of major importance to the study of evolution in the vertebrates. One is concerned with the origin of the frogs from more primitive amphibian types, while the other shows that a group of fishes from the Old Red Sandstone constitute a separate class of vertebrates, equal in rank to and ancestral to the remaining fishes.

Tracing, in this brilliant series of researches, the main stages of the descent of the mammals from their earliest fish-like ancestry, Professor Watson has certainly performed 'work of acknowledged distinction in the field in which Charles Darwin himself laboured'.

The BUCHANAN MEDAL is awarded to Sir WILSON JAMESON, formerly Dean of the London School of Hygiene and Tropical Medicine and since 1940 the Chief Medical Officer to the Ministry of Health and the Board of Education. In both capacities Jameson has shown himself to be a man of stimulating influence and leadership, determined and persistent in his efforts to ensure that advances of medical knowledge in the laboratory, the clinic and the field shall receive prompt application in administrative practice.

Largely to Jameson's vigorous policy is due the hope that active immunization against diphtheria, which has banished the disease from many large communities of North America, will at length find systematic and effective application in this country, where many of the discoveries were made which have rendered it safe and practicable. In the prompt official adoption of methods using modern technical resources, to deal with the recent increase of tuberculosis under war conditions, and in the recognition of adequate and scientifically planned nutrition of the people as a central item of an effective health policy, Jameson's active and enlightened influence can again be discerned.

Of the grounds on which the founders of the Buchanan Medal desired the awards of it to be made, Sir Wilson Jameson's high claim to it is based on 'administrative and constructive work' of outstanding merit in the service of Hygienic Science.

The HUGHES MEDAL is awarded to Professor ENRICO FERMI, now of New York. Professor Fermi has made most notable contributions both to theoretical and

experimental physics. In the early days of the modern quantum theory he was one of the first theoretical physicists to appreciate the generality of the considerations put forward by Pauli and known as the Exclusion Principle. This led him to discuss the statistical theory of a perfect gas of particles in equilibrium, obeying this principle, with results which were obtained independently and almost simultaneously by Dirac by similar methods. These results of Fermi and Dirac are of the utmost importance in the modern theory of assemblies of similar particles, such as electrons, protons, and neutrons. Following this outstanding personal contribution, Fermi played a great part in building up at Rome a distinguished school of theoretical physics, where he himself made one of the earliest successful attempts to construct a theory of radioactive β-ray change. This theory shows the most profound insight into the theoretical nature of the quantum theory.

His interest in the atomic nucleus led Fermi naturally on to his experimental studies in this field. Immediately after the discovery of the neutron he realized that it provided a new possibility of attack on the nucleus and of stimulating nuclear change by neutron bombardment. This work opened up the fruitful modern field of study concerned with the transformations of nuclei of medium and great atomic number, and led directly to the most exciting transformations of all, the nuclear fission of uranium and thorium.

Professor Fermi's work is characterized throughout by profound insight and great experimental skill. In the fields which he has made his own he is universally acclaimed a leader.

ISAAC NEWTON

from a portrait by Charles Jervas, in the possession of the Royal Society. This was presented by Sir Isaac Newton at the meeting held on 16 May 1717, and is recorded in the Journal Book under that date in the words: 'The President gave the Society his picture drawn by Mr Jervase for which he had their thanks'.

NEWTON
TERCENTENARY
CELEBRATIONS

AT THE ANNIVERSARY MEETING
OF THE ROYAL SOCIETY
30 NOVEMBER 1942

Address of the President
Sir Henry Dale, C.B.E.
Anniversary Meeting, 30 November 1942

We are to-day within a few weeks of the three hundredth anniversary of the birth of Isaac Newton. Wherever the progress of our Western science and philosophy has become effective, men will remember what that event was to mean for the world. Newton, as we shall hear, at the age of 43, when he had determined to abandon all further concern with natural philosophy, was induced at length, by Halley's friendly insistence, to give written form and system to the mathematical discoveries with which his amazing mind had been occupied over a period of some twenty years. The result was one of the greatest intellectual achievements in the history of mankind—the *Principia*, providing for more than two centuries a framework for the mechanical interpretation of the universe and a basis for the building of physical science, and therewith of the material structure of our modern civilization.

We in Britain regard Isaac Newton as still, beyond challenge, the greatest of our men of science. Nor should the claim be limited to this island or to the British Commonwealth of Nations; for it was not till nearly half a century after Newton's death that former British colonists in North America began their development of an independent nation; and Newton is theirs as well as ours.

But, while we may proudly claim him as the countryman of all who share the birthright of the English tongue, the discoveries of science have belonged, and must belong again, to the whole world, and Newton's achievement is a part of the common heritage of all peoples. It cannot be doubted that, if it had fallen in normal times, this tercentenary would have been marked by the greatest of international gatherings, in which men of science and philosophers from all the world would have assembled to do honour to Newton's memory. It would have been natural then to expect leadership, in such an enterprise, from the only two institutions which were intimately concerned with Newton's career as a man of science—Trinity College, in Cambridge, and this our Royal Society of London. Our two foundations did, indeed, confer as to the wisdom of attempting by joint action, even in this year of war, to arrange such a restricted and domestic celebration as the present conditions would allow. We agreed, however, to put aside such planning for the present, carrying it forward in our hopes to the time when a world at peace may be able to join in international commemoration of an event which has meant so much for all mankind. May the time be not too far distant.

To-day we are holding the 280th Anniversary Meeting of the Royal Society, on Saint Andrew's Day as by regular custom, ever since the first meeting on that day in 1662. It seemed to us that we should depart, on this occasion, a little from the

usual order of our proceedings, so that, on a day so near to the tercentenary of his birth, our Fellows and our guests may be reminded of Newton. We have accordingly asked three of our Fellows to address the Society on different aspects of Newton's work, in its relations to the science of the past and the present. We have asked Professor Andrade to give us the opportunity of understanding the magnitude of the change which Newton's work produced, in the conceptions of the material universe which were current in his own times. We know that to his contemporaries Newton's discoveries came as a great revelation, and we hope that Professor Andrade will help us to understand why they did so. We have asked Lord Rayleigh to deal, by demonstration where possible, with the experimental work of Newton and the great discoveries which he made by that method. This is an aspect of his greatness which popular estimates have tended to overlook; but I think that Lord Rayleigh will be able to convince us that Newton as an experimenter would have had claims to a place among our greatest men of science, even if he had failed, as he so nearly did fail, to write the *Principia*. And, finally, we have asked Sir James Jeans to give us some reassessment of the validity and permanence of Newton's system, in relation to the immense advances of knowledge in our own times. There are many who have not the mathematical equipment to follow them in detail, who are nevertheless aware that revolutionary changes have been taking place in conceptions of the mechanics of the universe and of its ultimate material units. How is the Newtonian system affected by the discoveries which have required the general theory of relativity and the quantum mechanics at opposite ends of the stupendous scale? Is it being supplemented, modified or superseded after its centuries of dominance? We hope that Sir James Jeans will tell us; and we may remember, perhaps with comfort, that Newton's *Principia* seemed difficult and abstruse to his contemporaries, and that he even confessed that he had made it so deliberately, 'to avoid being bated by little smatterers in mathematics'.

Before I call on our chosen lecturers to address us, there are two other matters relating to Newton and the Royal Society, which it seems proper to mention here. In the hamlet of Woolsthorpe, near Colsterworth on the Great North Road, some six miles south of Grantham, there is still a modest manor farm-house, with a small orchard in front of it. Here the Newtons lived, simple yeoman farmers, and here, two months after his father had died, Isaac Newton was born, a puny, premature infant, on Christmas day 1642, twenty years before the Royal Society was incorporated by the grant of its first Charter. The house stands but little altered since that day. The room in which Newton was born has a simple marble tablet on the wall, inscribed with Pope's well-known couplet. But this house had importance in Newton's later life and in his work, and not only as his birthplace. It was here that he returned from his schooling at Grantham, at the age of 16, to take charge of the farm for his mother; and here, to the incalculable gain of science and the world, he showed such incompetence as a farmer that he was sent back to school, and thence to Cambridge. It was here, again, that he returned in the autumn of 1665, when the plague drove him from Cambridge; and here, during the following

eighteen months of quiet exile in the country, his early ripening genius grasped already the essential principles of his major theoretical discoveries. You can still see the upper chamber which he then used as a study; and in the little orchard there is an old, recumbent apple tree, which, they will tell you, is descended by direct grafting from that which Newton saw. The land which Newton's family farmed is rapidly being laid waste, alas, by quarrying for ironstone; and soon there will be little left unspoiled save the orchard and garden round the house. It has seemed to us that, in this year of commemoration, something should be done to preserve for posterity a house and garden which carry such momentous memories, and which have meant so much for science. We have accordingly formed a small Committee, in which Sir John Russell and Sir James Jeans have joined with the Officers of the Royal Society. We have been in friendly negotiation with the lord of the manor, Major Turnor—a name associated in several past generations with the Royal Society and with Isaac Newton—as to the possibility of acquiring this now tiny but historic property, so that it may be put for as long as possible beyond the risk of damage or decay. Major Turnor has generously offered to sell the property for this purpose at a price substantially less than its value, and only this morning I received a letter from Lord Macmillan which enables me to announce that the Pilgrim Trust will be responsible for the sum required for the purchase.

Then I think that it is our special duty here, at this Anniversary Meeting, to remember that, while Newton's great discoveries belong to the world, they came to publication through the Royal Society, and that Newton occupied its presidential Chair for the last twenty-four years of his long life. Though his *Opticks* was not published till after he had become President, his original work for science was practically finished by the time of his election, and he had for some years been Master of the Mint. There can be no doubt, however, that the wide fame of his achievements, and the respect and admiration in which he was everywhere held, did much, at a critical period in its history, to establish the prestige of this Society in the eyes of the world. Let us then remember to-day that Isaac Newton, the greatest man of science of our race. was also the greatest of the Royal Society's presidents.

WOOLSTHORPE MANOR
from photographs taken by Walter Lee of Grantham, November 1942.

ISAAC NEWTON

from a portrait by Johan Vanderbank in the National Portrait Gallery.

Newton and the science of his age

By Professor E. N. da C. Andrade, F.R.S.

It is my task—my honourable and inspiring task—to say something of Isaac Newton as seen against the background of the science of his time. I shall try to display briefly the position as he found it and to resume in a small space his great achievements and the changes in outlook which they produced. In praising Newton I shall endeavour not to do injustice to his great forerunners and to the men of his time who pursued worthily the same great ends as he did, and who would have held the centre of the stage in any other age than that dominated by him. For Newton, like Shakespeare, did not stand as a lonely adventurer into new realms, though he travelled further and straighter than the rest. Shakespeare was the supreme poet and playwright at a time when poetry and plays were part of the life of every cultivated man and occupied the attention of the brightest intellects. Newton was the supreme scientist in an age when the quantitative method of questioning Nature was abroad in the air. Each was the child of his time.

Let us consider the position when Newton went to Cambridge in 1661. The hold of Aristotle, whose works had for centuries been the ultimate resort of all those seeking knowledge of the working of nature, had been shaken off by such men as Galileo and Gilbert, but most of the learned still thought that those who relied on experiment were pursuing a futile and impudent course. The first resolve of Marlowe's Faustus

> Having commenc'd, be a divine in shew,
> Yet level at the end of every art,
> And live and die in Aristotle's works

still represented the aim of many students. The foundation of the Royal Society in 1662 had been the occasion of many attacks on the experimental method, attacks stoutly met by Glanvill and by Sprat, and as late as 1692 Sir William Temple's *Essay upon the Ancient and Modern Learning*, satired by Swift in *The Battle of the Books*, set out to prove the superiority of the philosophers of the ancient world over all the moderns. Thus when Newton was a young man the new experimental method of questioning Nature was steadily making its way and the omniscience of the ancients was being called in doubt by a new school, but experimental science was by no means firmly established as a respectable study.

The great figures among the worthies of the exact sciences who had already appeared at that time were Copernicus, Tycho Brahe, Kepler, Gilbert, Galileo and Descartes. Kepler, following his great forerunners, had found the true laws of planetary motion, which were to be explained by Newton. Kepler's views as to the mechanism of the planetary motions were in his earlier writings largely mystical,

involving the perfect properties of the five regular solids and also certain motivating souls or spirits, *animae motrices* (figure 1). Throughout he held to the medieval point of view that a body could not maintain its motion unless there were a force propelling it. In his later writing he invoked a magnetic force, but
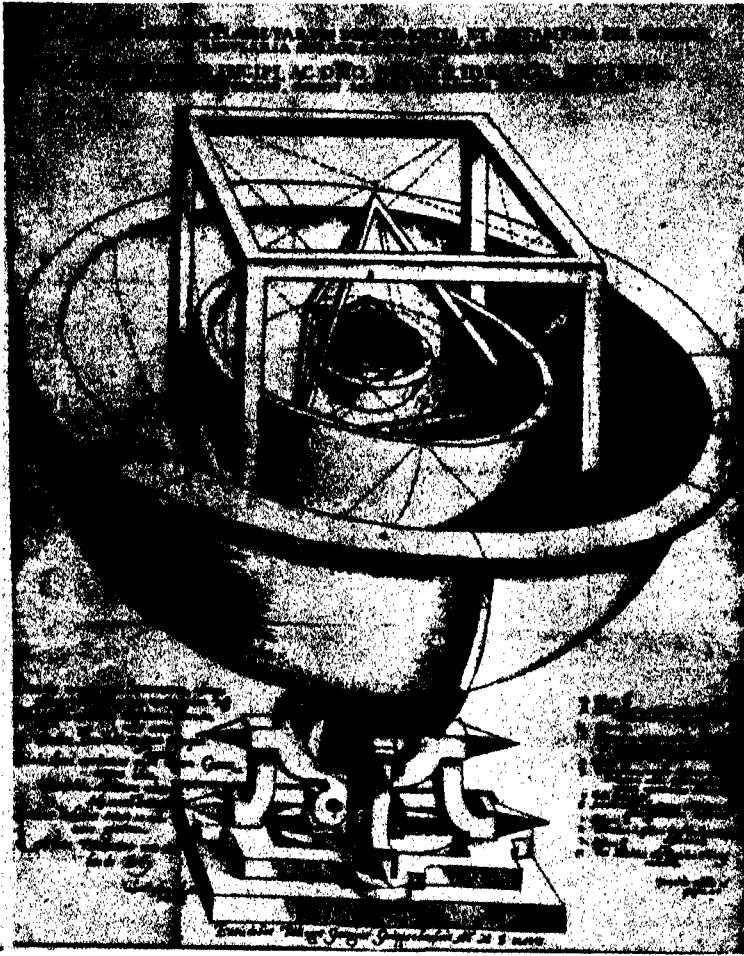
FIGURE 1. Illustration from Kepler's *Mysterium Cosmographicum*, 1596, showing the orbits of the planets fitted between the octahedron, icosahedron, dodecahedron, tetrahedron and cube.

it was not directed to the sun, like the true gravitational force, but pushed the planets on their way—*non est attractoria sed promotoria*. It was essentially bound up with the rotation of the sun. Thus he was ignorant of the basic laws of mechanics and his magnetic force had none of the true properties of magnetic forces. Kepler made no approach to a mechanical explanation of his laws.

Gilbert had not only established the basic principles of terrestrial magnetism and carried out fundamental work on electricity, but had invoked a force from the moon—a magnetic force it is true—to produce the tides. Galileo's greatest achievement had been to lay the foundations of mechanics. None of these men, however, had made any impression on the bulk of the learned: Francis Bacon, for instance, neglects Galileo and Kepler, and refuses to take Gilbert seriously. The great figure in the eye of natural philosophers was Descartes, who had developed a cosmogony based upon mechanical principles, not precise mechanical principles it is true, but principles very different from the mystical ones then in vogue. Joseph Glanvill used to lament that his friends had not sent him to Cambridge, where he might have learned the new philosophy of Descartes, rather than to Oxford, where Aristotelianism ruled. Descartes' system had acquired such a hold on men's minds that his views were still supported long after Newton's death. Both Jean Bernoulli, who died in 1748, and Fontenelle, who died in 1757, were Cartesians to the end. The demolition of the Cartesian system was, in contemporary eyes, one of Newton's greatest achievements. James Thomson said in his *Ode to the Memory of Sir Isaac Newton*, published immediately after his death:

> The heavens are all his own; from the wild rule
> Of whirling vortices, and circling spheres,
> To their first great simplicity restored.
> The schools astonished stood.

From the *Principia* it is clear that Newton himself derived particular satisfaction from having invalidated the Cartesian system. It is fitting, then, that we start our consideration of contemporary science by a glance at this system, not only because it was the only attempt before Newton to explain the motions of heavenly bodies on general principles, but because it furnishes a contrast which brings out the essential Newtonian point of view.

Descartes starts, in the spirit of medieval thinkers, from certain general philosophical principles. He decides that the fundamental property of matter is extension—impenetrability, colour, hardness and so on are only secondary characteristics. Extension, which has three directions, is the subject of mathematics: motion is the subject of mechanics. All the different qualities of different kinds of matter are provided by different motions of the minute parts of which it is composed. 'Give me extension and motion' declares Descartes, 'and I will construct the world.' One consequence of his fundamental belief is that there cannot be a vacuum, for extension without matter is a contradiction. He further considers, on theological grounds, that the quantity of motion in the heavens must be constant. He blames Galileo for founding his mechanics on experiments and not on reflexions on first causes. 'Everything Galileo says about the philosophy of bodies falling in empty space is built without foundation: he ought first to have determined the nature of weight.' Newton's point of view, of course, was the exact opposite to that of Descartes: he says in the famous letters to Bentley, '...for the cause of gravity is what I do not pretend to know, and therefore

would take some time to consider of it', and again, 'gravity must be caused by an agent acting constantly according to certain laws, but whether this agent be material or immaterial, I have left to the consideration of my readers'. For Newton, as for the best of his successors, science was concerned with the question of 'How?': Descartes, like the ancients, was concerned with the insoluble question of a fundamental 'Why?'.
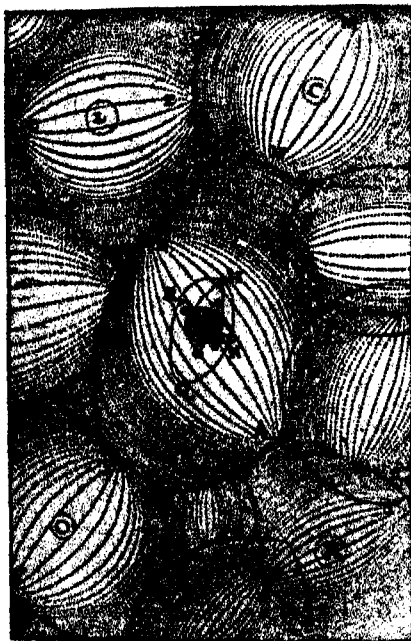
FIGURE 2. A page from Descartes' *Principia philosophiae*, 1644, showing in the middle the sun's vortex, governing the planets. Other vortices surround it.

It followed from the philosophic hypothesis of Descartes that the only kind of motion possible in a plenum was a motion in closed paths, more particularly a circular motion, since a particle could only move if another particle took its place. It was on grounds of this kind that he elaborated his vortex hypothesis. Certain very fine particles, which filled interplanetary space, moved round ceaselessly in huge vortices and carried the planets with them. The moon was carried round the earth by a minor vortex, and so on (figure 2).

Descartes' cosmogony, then, was founded on a philosophical system: it was pictorial and unquantitative. The paths of the comets, handed on from one vortex to another, were in particular irreconcilable with observation. There was no attempt to deduce Kepler's laws, or to show how anything but circular motion could result from the vortices. The whole scheme was spun from the brain of Descartes, with more or less casual references to actual phenomena. No doubt it was referring to Descartes that Roger Cotes said in his Preface to the second edition of the *Principia*, 'Those who fetch from hypothesis the foundation on which

they build their speculations may form indeed an ingenious romance, but a romance it will still be.' In spite of, or perhaps because of, this, Descartes' influence was immense, and when Newton was at Cambridge as a young man it was Descartes who was the great authority for all such progressive spirits as speculated on the structure of the universe.

We now turn to the astonishing story of the birth of the *Principia*. The time was ripe for the appearance of this great work. As regards the laws of motion, Galileo, whose services Newton freely acknowledges, and Descartes himself had done much to prepare the way for the more precise and particular formulation which Newton gives. Hooke among others had clearly expressed the protest of the most forward spirits of the time against the speculative method. 'The truth is, the Science of Nature has been already too long made only a work of the Brain and the Fancy: it is now high time that it should return to the plainness and soundness of *Observations* on *material* and *obvious* things.' The Royal Society was active in stimulating the pursuit of the new method and on the Continent the *Accademia del Cimento* had done excellent work, while the *Académie des Sciences* was founded in 1666. These continental societies had little influence on Newton, but are symptomatic of the general movement towards the experimental method. The stage was set for great things.

The story opens at Woolsthorpe, Newton's birthplace, whither Newton had come from Cambridge in June 1665 to escape the plague. He was then 22 years old and not yet a Master of Arts or Fellow of Trinity. He had read what he calls 'Schooten's *Miscellanies*' (probably the *Exercitationum Mathematicarum Libri V*), Descartes' *Geometria* and Wallis' works, and further was, of course, familiar with the work of his teacher Barrow. He had written his first treatise on the calculus, or 'fluxions' as he called it, but he had published nothing. The words which he wrote some fifty years later about this great springtime of his intellectual life have often been quoted but cannot well be omitted on an occasion like this. 'And the same year (1666) I began to think of gravity extending to the orb of the Moon, and having found out how to estimate the force with which a globe revolving within a sphere presses the surface of the sphere, from Kepler's Rule of the periodical times of the Planets being in a sesquialterate proportion of their distances from the centers of their orbs I deduced that the forces which keep the Planets in their Orbs must [be] reciprocally as the squares of their distances from the centers about which they revolve: and thereby compared the force requisite to keep the Moon in her orb with the force of gravity at the surface of the earth, and found them answer pretty nearly. All this was in the two plague years of 1665 and 1666, for in those days I was in the prime of my age for invention, and minded Mathematicks and Philosophy more than at any time since. What Mr Hugens has published since about centrifugal forces I suppose he had before me.'

It seems likely that he had already had his laws of motion in his head—in any case it is clear that he was convinced that every body would continue to move uniformly in a straight line unless some force acted on it, and that, therefore,

there must be some force acting on the moon which drew it away from the straight line, tangential at any moment to its path, in the direction of the earth. Treating the moon's path as circular, from Kepler's third law, and from the law connecting the centrifugal force with the radius and the velocity, or the equivalent proposition to which Newton refers, it is easy to deduce the inverse square law. To show that the force keeping the moon in her orbit is the earth's gravitational force, assumed to diminish as the inverse square, is, however, a further step demanding a computation of how strong the gravitational force at the moon's orbit will be, compared to the measured force at the surface of the earth. Newton made this step and found it 'answer pretty nearly'. Why, then, did he delay the announcement of the law of gravity for twenty years or so?

There is always a ready answer to questions of this kind where Newton is concerned—that he never published anything until invited, in general strongly urged, to do so. It is, however, clear from many signs that Newton was not himself satisfied about the matter until some time about 1686. The usual story is that he took a wrong radius for the earth, namely one corresponding to 60 miles for 1° of latitude instead of the correct value of about 70, but this story is very improbable on many grounds, one of which is that good values were readily available to him.

The real reason for Newton putting the work aside seems to have been that the calculation, as far as the force at the earth's surface is concerned, depends essentially upon it being legitimate to assume that the earth's mass may be considered as concentrated at the centre. That this assumption is valid is far from obvious. It is fairly clear from certain passages in *De Motu* and in the *Principia* that it gave Newton some trouble to prove this assumption and that he did it late. In a letter to Halley of 20 June 1686 he says, 'I never extended the duplicate proportion lower than to the superficies of the earth, and before a certain demonstration I found the last year, have suspected that it did not reach accurately enough down so low.' Although this refers to the gravitational force within a sphere, this and the point under discussion are involved in the same mathematical demonstration, which Newton gives in the *Principia* in Book I, proposition LXXI and other propositions following it. In any case Newton appears not to have been satisfied with his first calculations and to have turned to other things, possibly his optical experiments.

It is a strange thing that the *Principia* owes its publication largely to a quarrel with Hooke, and its sequel. In 1679 Hooke, then acting as Secretary of the Royal Society, wrote to Newton about various scientific matters and asked him very civilly for a philosophic communication—a paper, as we should say nowadays. He also asked for Newton's opinion on his *Potentia Restitutiva* and 'particularly if you will let me know your thoughts of that [hypothesis] of compounding the celestial motions of the planets of a direct motion by the tangent and an attractive motion towards the central body'. In his reply Newton made an extraordinary remark. 'But yet my affection to philosophy being worn out, so that I am almost as little concerned about it as one tradesman uses to be about another man's trade or

a countryman about learning, I must acknowledge myself averse from spending that time in writing about it which I think I can spend otherwise more to my own content and the good of others: and I hope neither you nor any body else will blame for this averseness.' This is but one of many occasions on which Newton expresses his disinclination, almost distaste, for any further scientific work, his first antipathy having been aroused by the disputes and misunderstandings consequent on the publication of his first great paper on the prism.

To return, Newton did comply with Hooke's request for something for the Society by pointing out that a body let fall from on high should strike the earth slightly east of the perpendicular, and gave precise and excellent directions for carrying out the experiment. The purpose was to prove the diurnal rotation of the earth. Hooke, in reply, pointed out that the ball should fall to the south as well as to the east, and further corrected Newton in a point which is too complicated to discuss here and one on which differences of statement can be due to different interpretations of the problem. This correction, tactlessly expressed, irritated Newton in the highest degree and he answered curtly. In further letters, written in apparent unconsciousness of the annoyance he had given, Hooke suggested that the law needed to explain the planetary motions was the inverse square law.

Now other men had come to the same conclusion. In particular, Wren and Halley had discussed with Hooke the possibility of explaining the mechanism of the heavens on the basis of an inverse square law. Hooke declared that he could demonstrate mathematically that the path of a particle in a central inverse square field would be an ellipse, but it is clear that he never did so, and equally clear that he had not the mathematical equipment necessary to begin an attack on the problem. He was in the unfortunate position of being entirely convinced of a truth that he could not prove. In August 1684 Halley visited Cambridge and asked Newton what the path would be. He replied that it would be an ellipse and that he had formerly calculated it. He could not find the calculation but soon sent a proof (or apparently two different proofs) to Halley.

This incident seems to have aroused Newton from the distaste for science into which he had fallen, and he put together the treatise *De Motu*, founded on a course of lectures, which Halley presented to the Royal Society on 10 December 1684. The story of how Halley then coaxed and cajoled Newton into writing the *Principia* is familiar, but we in this Society ought not ever to celebrate the great work without a tribute to Halley, who not only realized at once the fundamental importance and significance of Newton's work but used all his tact to get the book written and made himself financially responsible for the production, the Royal Society being in financial difficulties at that time. We are not now in financial straits and shall, I believe, be glad to bear the expense of producing a second *Principia* when the genius of our age brings it forth. The book appeared in 1687, and bears the imprimatur of the then President (figure 3), who, although he achieved nothing in science, is still remembered. He was Samuel Pepys.

The *Principia* is not an easy book to read. The proofs are all given in form of classical geometry, although, since it is certain that at the time when it was written Newton was in possession of the fundamental processes of the calculus and of the methods of analytical geometry, it is unlikely that this was the form in which he first derived them. Figure 4 shows a typical geometrical diagram, dealing with the moon's motion, and a comparison with the pictures from Kepler and Descartes gives a graphic representation of the changed spirit in handling the mechanism of the heavens. With reference to the mathematical methods of the
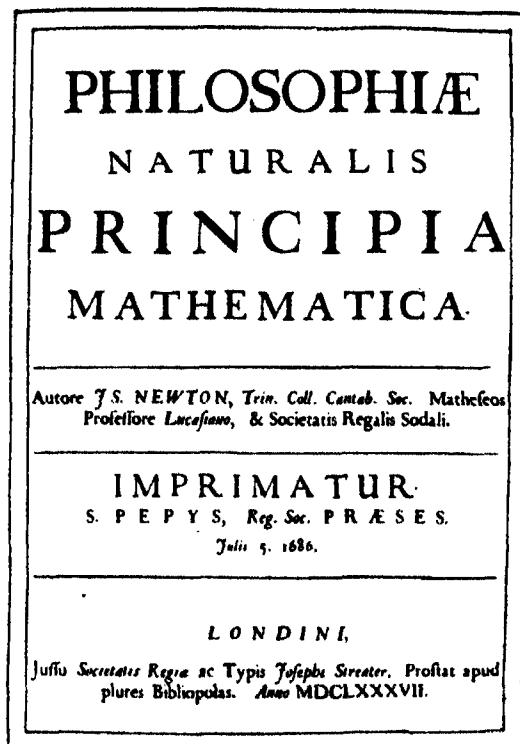
FIGURE 3. Reduced facsimile of the title page of the *Principia*,
bearing the imprimatur of Samuel Pepys.

Principia Whewell has said 'Nobody since Newton has been able to use geometrical methods to the same extent for the like purposes; and as we read the *Principia* we feel as when we are in an ancient armoury where the weapons are of gigantic size; and as we look at them we marvel what manner of man he was who could use as a weapon what we can scarcely lift as a burden.' Various conjectures have been made by Rosenberger, Cantor, Giesel, Gerhardt and others as to why he did not use the new methods: Rouse Ball thinks that he was probably unwilling to add to the difficulties by introducing a new mathematical method. Be that as it may, he had a horror of unfounded criticism, and, as he told Dr Derham 'to avoid being baited by little smatterers in mathematics,

he designedly made his *Principia* abstruse; but yet so as to be understood by able mathematicians'. That he did not underrate the difficulty of the work is clear from what he says in the beginning of Book III: 'I chose to reduce the substance of that book into the form of propositions (in the mathematical way) which should be read by those only, who had first made themselves masters of the principles establish'd in the preceding books. Nor would I advise any one to the previous study of every proposition of those books. For they abound with such as might cost too much time, even to readers of good mathematical learning. It is enough if one carefully reads the definitions, the laws of motion, and the first three sections of the first book.'*
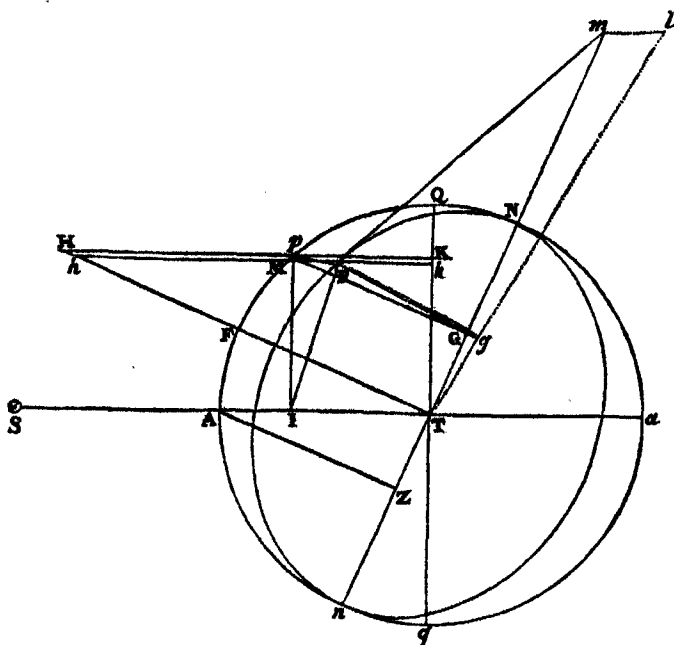
FIGURE 4. A typical geometrical diagram from the *Principia*.

The first book contains certain definitions of space, mass and time which have often afforded a theme ample enough for discussion. But we will here let them pass uncommented. The famous laws of motion owe much to the labours of previous workers, in particular to Galileo, whose services Newton clearly acknowledges. The simple laws of central orbits under an inverse square law are worked out in much detail and the laws of pendulum motion are developed, with due acknowledgments to Huygens, who had published his *Horologium Oscillatorum* in 1673.

In the first book the motions are all supposed to take place in a non-resistant medium. In the second Newton considers motions in a resisting medium, always with his eye on the Cartesian world system which he was intent to demolish. He deals with a resistance proportional to the velocity and with a resistance proportional

* Andrew Motte's translation of 1729, volume 2, page 201.

to the square of the velocity, and further points out different kinds of fluid resistance, which he later defines most clearly in the 28th Query appended to the third edition of the *Opticks*, 'for the resisting Power of the fluid Medium arises partly from the Attrition of the Parts of the Medium and partly from the *Vis inertiae* of the Matter', that is, partly from the viscosity and partly from the bulk motion of the medium. The latter resistance he assumes proportional to the square of the velocity. In this book he opens the way to the hydrodynamics of real fluids. In the part dealing with hydrostatics he proves the law of the diminution of atmospheric pressure with height. He then further discusses the motion of the pendulum and is the first to suggest its use for making a survey of the gravitational acceleration. In another connexion he derives an expression for the velocity of sound, this being the first case of the calculation of the velocity of a wave from the properties of the medium. The only other point in this book to which I will refer is the calculation which Newton carries out on the motion of an infinitely extended viscous fluid in which a body rotating upon its axis is immersed. What he has in mind is the Cartesian vortex—'I have endeavoured in this proposition to investigate the properties of vortices, that I may find whether a celestial phenomenon can be explained by them.' He effectively defines the force 'arising from the want of lubricity in the parts of the fluid' as proportional to the velocity gradient, whence the term 'Newtonian viscosity'. He is thus the first to touch the mathematics of viscous fluids. He finds that the periodic time of circulation of the fluid carried round by a rotating sphere is proportional to the square of the distance from the centre of the sphere,* which is grossly inconsistent with Kepler's third law. Newton considers that his deduction is a clear refutation of the Cartesian vortices and, for once, almost gloats over his victory. 'Let philosophers then see how the phenomenon of the sesquiplicate ratio can be accounted for by vortices.' He brings many other objections against the Cartesian vortices: he points out, for instance, that a continuous supply of energy will have to be given to the sphere to maintain the motion, because 'it is plain that the motion is proportionally transferred from the centre to the circumference of the vortex, till it is quite swallowed up and lost in the boundless extent of that circumference'.

The second book of the *Principia* is the foundation stone of mathematical physics. In it Newton shows an extraordinary instinct for grasping the essentials of a problem—it would almost seem that he knew the solution in advance and added the proof as a concession to those less clear-sighted. As William Whiston, who knew him well, and succeeded him in the Lucasian chair, says: 'Sir Isaac, in mathematics, could sometimes see almost by intuition, even without demonstration....And when he did but propose conjectures in natural philosophy, he almost always knew them to be true at the same time.' The manner of his refutation of the Cartesian vortices introduces a new spirit into the discussion: no hypothesis

---

* Actually this is an error, which appears never to have been pointed out, although Stokes has indicated the like error in the case of the rotating cylinder, which Newton works out. It should be the *cube* of the distance.

about the heavens is tenable unless the quantitative deductions from it agree with observation.

The third book opens with an introduction where, after stating what has been done in the first two books, Newton sets down the superb sentence—'Superstat ut ex iisdem principiis doceamus constitutionem systematis mundani'—it remains that from the same principles we demonstrate the form of the system of the world. It is this third book which based celestial mechanics so firmly that what was done in the next two hundred years was rather extension of, and improvements on, the Newtonian method than anything radically new. Not only does Newton establish the movements of the satellites of Jupiter, Saturn and the Earth, and of the planets round the Sun (or rather, as he points out, round the centre of gravity of the solar system) in terms of his gravitational theory, but he shows how to find the masses of the sun and planets in terms of the earth's mass, which he estimates quite closely; he accounts for the flattened shape of the earth and other planets; calculates the general variations of g over the surface of the earth; explains the precession of the equinoxes by consideration of the non-sphericity of the earth; calculates the main irregularities of the motion of the moon and of other satellites from the perturbing effect of the sun; explains the general features of the tides; and finally treats the orbits of comets in a way that shows that they are members of the solar system and enables the return of Halley's comet in 1759 to be accurately calculated. This brief and imperfect catalogue is merely a reminder of the scope of this extraordinary book, which drew from Laplace, no enthusiast, '... all this, presented with much elegance, assures to the *Principia* preeminence over all the other productions of the human mind'. The book closes with the famous General Scholium which returns to the confutation of Descartes' vortices and says, concerning the cause of gravity, 'Hypotheses non fingo'.

We have a portrait of Newton, by Kneller, at about the time of the publication of the *Principia* which is particularly impressive. In contrast to most of the later and more formal portraits we see him in his own hair and in the casual clothing which we may suppose him to have worn when at work. The look of wild, almost hostile remoteness and of dominating and piercing intelligence seem to show that the artist has well read the features of his sitter and given us a true picture of the man in his hours of creative thought.

Although the book was eagerly bought, the Newtonian method and discoveries made way but slowly. Biot says that of Newton's contemporaries three or four only were capable of understanding the *Principia*, that Huygens only half adopted the ideas, Leibniz and Jean Bernoulli fought against them, and that fifty years had to pass before the great truth demonstrated by Newton was understood by the generality of men of science, let alone developed. No doubt the difficulty of the book had much to do with the tardy appreciation, outside a narrow circle, of its contents. The Cartesian scheme was easy, pictorial, general: the Newtonian difficult, mathematical, precise. The very method of attacking the problem was altogether new. The second edition appeared in 1713, edited by Roger Cotes, of

whom Newton said 'if Mr Cotes had lived we might have known something'. The preface clearly shows that among the learned the Newtonian scheme had not been widely accepted. On the Continent the progress was still slower and it is generally held that it was Voltaire's *Elémens de la Philosophie de Newton*, which appeared in 1738, that led to Newton's work being appreciated in France. Incidentally, the story of the apple, which Voltaire had from Newton's niece, Mrs Conduitt, appeared in the second edition of Voltaire's book, in 1741. Later, it was in France that Newton's work was raised to great glory, when Lagrange and Laplace erected edifices of splendour and elegance on Newtonian foundations.

The optical work of Newton was perhaps his favourite study: at any rate it was that to which he made what is probably the only enthusiastic reference which he ever permitted himself, when he said, of his fundamental discovery in the matter of prismatic colours, 'being in my judgment the oddest, if not the most considerable detection which hath hitherto been made in the operations of nature'. The beautiful experiments which formed the background of this work will be dealt with by Lord Rayleigh, so I shall say nothing of them, but a few words as to the Newtonian attitude and theory may be permitted.

The nature of colour had from the days of Aristotle been the subject of philosophical speculation. According to Aristotle, colours are a mixture of light and darkness, or of white and black, a view which, embellished and modified in various ways, survived Newton and appeared again in, for instance, Goethe's writings. Descartes was apparently the first to break really new ground in comparing colours to notes in music: his view of light was that it was a pressure transmitted through the particles that filled all space, and he conjectured that a rotation of the particles might be the effective cause of colour, a view which presented inherent difficulties effectively exposed by Hooke. We may agree with Huygens that 'Descartes has said nothing that is not full of difficulties or even inconceivable, in dealing with light and its properties'. The medieval question as to whether light was a substance or an accident was still occupying the attention of even acute experimentalists such as Grimaldi, the discoverer of diffraction, whose book appeared in 1665. It was Hooke again, who, with his extraordinary flair for the truth, combined with an inability to overcome the last difficulties that stood between him and a convincing conclusion, initiated the modern views. In his *Micrographia*, 1665, he expressed the view that light was a very quick vibration propagated with a finite velocity. He gave what resembles Huygens' construction for finding the wave front on refraction, but, since he thought that light travelled faster in a solid medium, e.g. glass, than in air, he found that in the medium the wave front must make an acute angle with the ray. It was with this 'obliquity' that he connected colour. 'Blue is an impression on the Retina of an oblique and confus'd pulse of light, whose weakest part precedes, and whose strongest follows. ...Red is an impression on the Retina of an oblique and confus'd pulse of light, whose strongest part precedes and whose weakest follows.' For him blue and red were the primary colours, all others being mixed. There is no time to follow his

extraordinarily acute experiments on the colours of thin plates, but it must be noted that his ill-tempered attacks were the cause of Newton expressing a disgust with science which nearly caused him to abandon her pursuit, and were, it is almost certain, the reason why the *Opticks* was not published until 1704, the year following Hooke's death. The quarrel between the two men, both, as is evident from their correspondence, capable of generous appreciation of the other's achievements, was exacerbated by Oldenberg, then Secretary of the Society, whose dislike of Hooke may have sprung from causes little creditable to him. Newton had been undoubtedly stimulated by his reading of Hooke's *Micrographia*, and he was always very remiss in acknowledgements to him. Let us pay a tribute to poor Hooke, sickly and without position or powerful friends. He had not Newton's power of thought, but he was probably the most ingenious contriver who ever lived and was a shrewd and daring speculator.

The experiments on the composition of white light, which reduced the whole study to a quantitative basis, by showing that the refraction is a measure and index of the colour, will, as I have said, be dealt with by Lord Rayleigh. To the reflecting telescope, the explanation of the rainbow, the work on thin films, exemplified by Newton's rings, and the experiments on diffraction, this passing reference must suffice.

Nowhere more than in his writings on light does Newton stress his dislike of speculation not firmly rooted in experiment. The first words of the *Opticks* are 'My design in this Book is not to explain the Properties of Light by Hypotheses, but to propose and prove them by Reason and Experiments,' and again, in the 31st Query appended to the third edition of that book he makes his standpoint very clear, saying 'These Principles I consider not as occult Qualities, supposed to result from the specifick Forms of Things, but as general Laws of Nature, by which the Things themselves are form'd: their Truth appearing to us by Phaenomena, though their Causes be not yet discover'd. For these are manifest Qualities, and their Causes only are occult. And the *Aristotelians* gave the Name of occult Qualities not to manifest Qualities, but to such Qualities only as they supposed to lie hid in Bodies, and to be the unknown Causes of manifest Effects....To tell us that every Species of Things is endow'd with an occult specifick Quality by which it acts and produces manifest Effects, is to tell us nothing.' It is in the light of this that we must read the famous 'Hypotheses non fingo'. Newton, of course, did make hypotheses, and even call them such—for instance in *Principia*, Book II, Section IX, his assumption as to the behaviour of viscous fluids is headed 'Hypothesis' and there are other instances. All he meant was that he was reluctant to speculate beyond any possibility of quantitative deduction, to form conjectures whose defence would be merely a matter of dialectics.

Newton's corpuscular hypothesis to account for the experimental behaviour of light is clearly a hypothesis, but we shall see how closely he adapts it to the observations. First, he points out that light cannot be a wave motion, or it would spread out on passing through an opening. Figure 5 is the diagram with which

he illustrates this point. He knew, it is true, the phenomena of diffraction, but he did not realize how by making the wave-length small enough this could be reconciled with the general facts of rectilinear propagation, for which a stream of particles seems best suited. He clearly saw that the phenomena of the colours of thin plates demanded a periodicity, and he introduced this into his theory by the hypothesis of fits of easy reflexion and easy transmission. It is clear that
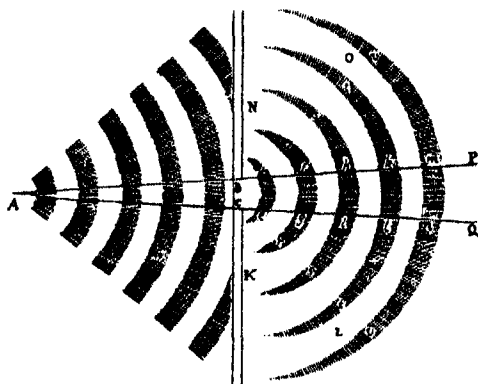

FIGURE 5

light is partly reflected and partly transmitted at the surface of a transparent body: Newton supposed that a light particle alternated at regular intervals be-tween a state in which it was transmitted through the surface and one in which it was sent back. He put forward diffidently ('Those that are averse to assenting to any new Discoveries but such as they can explain by an Hypothesis, may for the present suppose') the idea that the impact of the particle on the surface excited vibrations in the medium which, overtaking the particles, put them into these alternating states. We are forcibly reminded of modern theories. The length of the interval of the fit, corresponding to our wave-length, was greater for the red than for the blue, and Newton gives the interval for yellow light incident normally as just about the actual wave-length of yellow light. However, of his whole theory he says 'But whether this Hypothesis be true or false I do not here consider. I content myself with the bare Discovery that the Rays of Light are by some cause or other alternately disposed to be reflected or refracted for many vicissitudes.' When he comes to consider polarization, he has to endow his par-ticles with 'sides' so that there is a lack of complete symmetry about the direction of propagation. In short, he gave his particles just those properties which interpreted the experiment and hence was led to endow them with a periodicity and a polarity. This brought him to assume subsidiary waves accompanying the particles when they interacted with matter.

There seems to be a general belief that Huygens, as against Newton, advocated something very close to the wave theory of light, as it was accepted in, let us say, 1900. Huygens' wavelets, however, lack the essential properties with which

Fresnel afterwards endowed them: they were not only longitudinal but had what were practically particle properties, as exemplified by the fact that the pole—the point of contact with the envelope—alone was efficacious. He was just as incapable as Newton of giving a satisfactory explanation of diffraction.

If we are asked to state in a sentence what was the main effect of Newton's work on the thought of his time, I think that the answer must be that it was to establish the power and universality of the methods of quantitative science. To Galileo we owe the great service, one that cannot be too highly praised, of having made particle dynamics into a science, but he did not look beyond the earth for its efficacy nor suggest the application of his methods to the relative movement of the parts of a continuous medium, such as water. Huygens founded the study of rigid dynamics. Hooke suggested and speculated with extraordinary ingenuity and acuteness. Newton, however, showed that three clearly enunciated laws of motion applied to all observable movements of inanimate nature: they governed the motion of waves and projectiles, visible solids and invisible air, resisted as well as free movements. Together with the inverse square law they explained not only the gross movement of planets and the movement of the comets, which before had seemed capricious, but also details which nobody before had ever considered as being mechanically explicable, such as the precession of the equinoxes. The problems of the tides and of the irregularities of the moon's motion he did not fully solve, it is true, but he did enough to convince mathematicians that they were soluble by his methods. After Newton's work had been assimilated, the body of natural philosophers accepted it as a commonplace that all terrestrial and celestial movements were explicable in precise and numerical terms by calculations based on a few general laws: before Newton most thinkers were ready to invoke *ad hoc* principles and occult causes, based on human and divine analogies, for any but the simplest terrestrial phenomena, and the few who were in advance of their times were feeling tentatively for solutions which eluded their grasp.

Even in chemistry Newton was looking for an explanation in terms of attractions, though, strangely enough, in this science he never seems to have applied his own rule and made quantitative experiments. His work on light lies somewhat outside the mechanical scheme, but here again his insistence on the quantitative created a completely new attitude towards colour. It became a subject for measurement and calculation, rather than one for discussion in terms of generalities.

If we are to try to represent Newton's achievements by some modern analogy, to construct some imaginary figure who should be to our times what Newton was to his, we must credit this synthetic representative with, I think, the whole of relativity up to, and somewhat further than, the stage at present reached—we must suppose our modern Newton to have satisfactorily completed a unitary field theory. In light we must credit him both with having established the existence of spectral regularities and with their explanation in terms of the quantum theory. Possibly, too, we must give him the Rutherford atom model and its theoretical development, a simple astronomy in little to correspond to the solar system. Let us, then,

think of one man who, starting in 1900, say, had done the fundamental work of Einstein, Planck, Bohr and Schroedinger, and much of that of Rutherford, Alfred Fowler and Paschen, say, by 1930, and had then become, say, Governor of the Bank of England, besides writing two books of Hibbert lectures and spending much of his time on psychical research, to correspond with Newton's theological and mystical interests. Let such a man represent our modern Newton and think how we should regard him. Only so, I think, can we see Newton as he appeared to his contemporaries at the end of his life.

There are no discontinuities in nature and there are none in the history of science. No discovery or fundamental innovation is absolutely new, unconnected with past thought and the stirring spirit of its own time. Newton was not uninfluenced by certain of his immediate predecessors and of his contemporaries. The revolt from the introspective method of constructing explanations of heavenly and earthly phenomena by appeals to philosophic necessity had begun before his birth, and his time was rich in brilliant exponents of the experimental philosophy, whose names will always stand as stars adorning the story of science. To compare him with other men of his time and to recognize their contributions to the development of the physical sciences does not, however, lead us to think less of Newton's achievements but rather to wonder at them all the more. It is easier to estimate the size of a colossus if there are statues of more than life size in its neighbourhood than if it stands alone in a desert.

Newton owed much to the pioneer labours of Galileo, who had founded the science of mechanics, and, in a different way, much to Barrow, whose great mathematical acuteness and sympathetic support were always at the disposal of the young Cambridge scholar. To Hooke he owed more than he was ever prepared to acknowledge. From his great continental contemporaries, Huygens and Leibniz, he borrowed little, if anything. Possibly if mathematical advances were alone in question Leibniz would have to be considered as a possible rival, but for Newton mathematics were merely a means to a physical end: his mathematical innovations may even be left out of consideration without grave injury to Newton's fame. There is no record of physical experimentation that can compare for mastery and elegance with the *Opticks*, no work in exact science that produces the impression of supreme greatness and power of thought that the *Principia* does.

The spirit of this age is a denigrating one, which, in a reaction from the earlier custom of representing great men as free from all blemish and weakness, delights in attempting to show that nobody much exceeds the common level. If an earlier painter might have left out Cromwell's warts, a painter of to-day might represent his face as one huge wart. I have read record of Newton's weaknesses and I know of the adumbrations of his discoveries that can be found in forerunners and contemporaries: I acknowledge that his earlier biographer shut his eyes to any incident, writing or action that might seem to detract from his perfection. Nevertheless, all things considered, I think that the contemporary judgement of his greatness can still stand, and that, if the Marquis de l'Hôpital's query as to
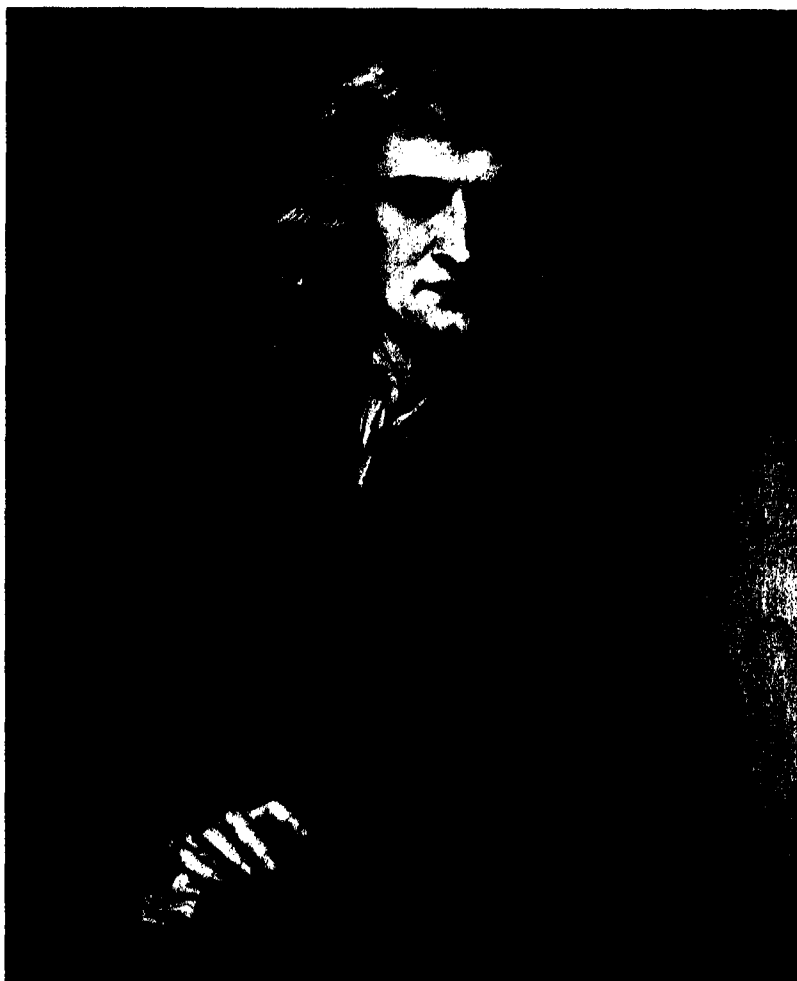
ISAAC NEWTON
from a portrait by Godfrey Kneller in the possession of the Earl of Portsmouth.
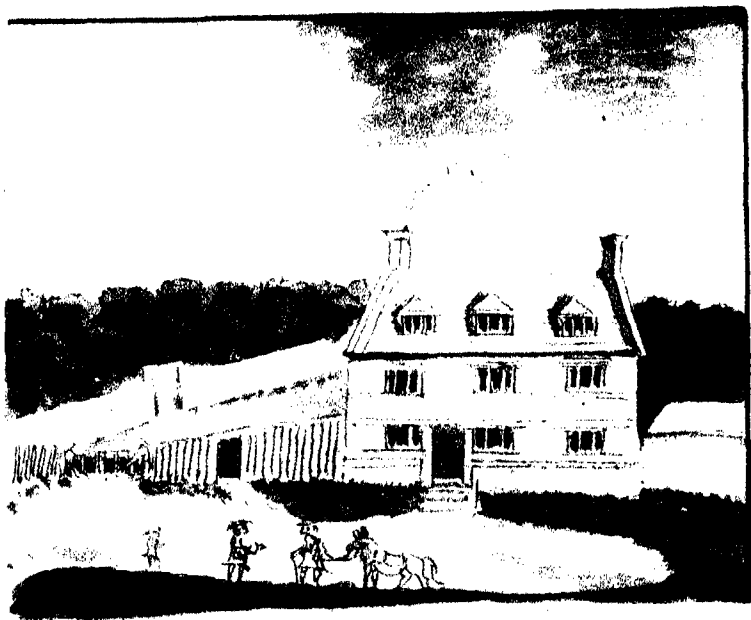
The Manner house of Wulsthorp in the parish of Collerworth Lincolnshire, where Sr. Isac Newton was born; being his own estate.

WOOLSTHORPE MANOR

from a coloured sketch by William Stukeley, F.R.S., in his *Memoirs of Sir Isaac Newton*, a manuscript book in the possession of the Royal Society, presented by A. Hastings White, lately librarian of the Society.

whether Newton eat drank or slept like ordinary men ('for I picture him to myself as a celestial genius) seems to our present-day sobriety an affectation, nevertheless we may agree that the line of Lucretius placed on the Trinity statue was well chosen and fitting—

Qui genus humanum ingenio superavit

who excelled the human race in power of thought

The death mask

Newton as an experimenter

By Lord Rayleigh, F.R.S.

The duty has been assigned to me of telling you something about Newton as an experimentalist. As the result of a study of what is known of his history, it seems to me that among his various intellectual pursuits experiment was his first love and the love to which he was most constant. Strange though it be, he seems in some moods to have doubted whether his theoretical studies were worth while, and I do not recall any case where he expressed himself enthusiastically about them. On the other hand, he speaks of his optical work as 'The oddest if not the most considerable detection which has hitherto been made in the operation of nature.'

Newton loved the mechanical side of experimental work. As a boy he constructed sundials, and, what is more, fixed one of them into the side of the house effectually enough for it to be there a century later. A notebook of his boyhood shows him assiduous in collecting recipes for various kinds of drawing materials, and he notes methods of performing some (rather nasty) conjuring tricks. Later on, when he is making his reflecting telescope, it is obvious that he is a skilled amateur mechanic, at home in furnace operation. He builds his own brick furnace, prepares speculum metal, and is apparently more successful than the professional opticians of the time in grinding and polishing it to a satisfactory spherical figure. (The days of parabolizing were not yet.) It was not until a good many years later that they were able to put such instruments on the market.*

Asked in his old age where he got the tools for his work, he replied that he had made them himself, and could have achieved little progress without doing so.

There are occasional hints to be gleaned that Newton practised other mechanical arts. Thus, when he examines the colour of thin blown glass, it appears that he has the facilities for glass-blowing at hand, and was presumably able to use them.

So much for the base mechanical side. Newton, however, had what may be called the itch of experiment and instinctively examined in this way any natural phenomenon that excited his interest. This instinct is not a common one, and it would be of interest to investigate statistically whether it is more correlated with mathematical aptitude than with (say) an aptitude for literary and historical studies. (Newton had all these.) Although he experimented in other fields, such as mechanics, heat, and electricity, and even in anatomy and physiology, his optical experiments are of much greater importance, and in the short time at our disposal

* Among the first successful commercial makers of reflecting telescopes was James Short (1710–1768), whose instruments were of the Gregorian type. He is said to have made a considerable number of concave and convex mirrors, and to have 'married' them by trial of what pairs gave the best result.

we shall only be able to consider a part even of these. The fundamental researches on the composition of white light were read before this Society in 1672, and subsequently recapitulated in his *Opticks* (1704), and we cannot do better than concentrate our attention on them. Although the results are common property nowadays, yet on an occasion like this we shall do well to go back to Newton's own methods and point of view, and to repeat his experiments as nearly as we can in his own way.

Newton's experiments on the spectrum are sometimes presented as if he had started out with the idea of examining the composition of white light. It is true that his *Opticks* (like his *Principia*) introduces the various topics as Theorems or Problems proposed after the manner of Euclid, and most of the experiments which I hope to show you are ranged in support of his proposition that 'The light of the sun consists of rays differently refrangible'. It does not seem probable, however, that he set out in the first instance to prove this or any other proposition. He bought a prism at Stourbridge Fair (near Cambridge) in 1666, 'to try therewith the celebrated phenomena of colours'. It is clear from this that the prismatic colours were quite a well-recognized phenomenon at this time, and this is also shown by the circumstance that the art of cutting diamonds so as to display them was already long known. In Peacham's *Gentlemanly Exercises* (1612) reference is made to 'A three square cristal prisme wherein you shall perceive the blew to be outmost next to the red,' and Grimaldi and others had already experimented on the subject, though without arriving at clear views. We can readily imagine how Newton, handling the prism, would soon find that the colours were well seen in candle light, but not in diffused daylight. It would not be a long step from this to try the effect on a beam of direct sunlight admitted through a hole in the shutter.

There is no reason to think that he did this with a very clear anticipation of what the effect would be. He was exploring a nearly virgin territory. His beam of sunlight passing through the hole produced an image of the sun. It was what we now call a pinhole image, though the hole need not be very small. Newton's hole was ⅓ in. diameter. Then he interposed the prism (figure 1).
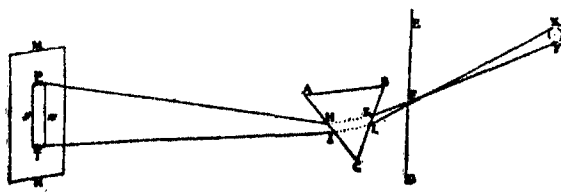

<center>FIGURE 1</center>

In spite of our increased control over nature, we cannot produce a beam of sunlight at pleasure late on a November afternoon, so I must be content to use the electric arc with a positive carbon presented end on, and the incandescent positive crater will represent the sun's disk. I have this in a lantern and a metal

diaphragm is placed some 2 ft. in front. (No lenses are used.) You see the image of the crater on the screen as a round disk. We will now do what Newton did, and place a glass prism in front of the hole. You see that the beam is deflected through a considerable angle towards the thick end of the prism, and you see that the image now takes the form of a band of prismatic colours on the screen. We could make this experiment more spectacular with modern resources, but my object is to follow in Newton's footsteps as nearly as circumstances allow. To most people the colours would seem the main feature of this experiment, but Newton was more surprised and impressed with the fact that the sun's image was no longer round, but was spread out into a band of which the length was some five times the breadth. The breadth of the image as projected through the prism still answered to the sun's diameter, that is to say, it was equal to the diameter of the original round image of the sun at the same distance with the prism away.

'If', he says, 'the refraction were done regularly according to one certain proportion of the sines of incidence and refraction, as is vulgarly supposed, the refracted image ought to have appeared round.'

At the same time the colours had to be taken into account. Were the colours definitely related to the differences of refraction, or was the apparent relation only incidental? In Newton's words:

'Wherever this inequality arises, whether it be that some of the incident rays are refracted more, and others less, constantly or by chance, or that one and the same ray is by refraction disturbed, shattered, dilated and as it were split and spread into many diverging rays, as Grimaldi supposes, does not yet appear by these experiments, but will appear by those that follow.'
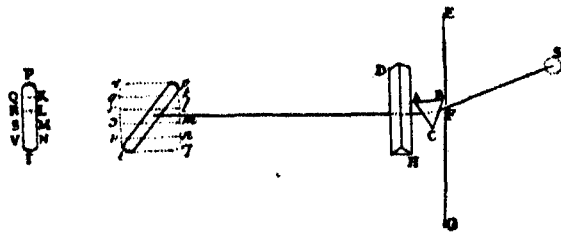
FIGURE 2

The next experiment was accordingly designed to throw light on this point. It is known as the experiment of crossed spectra (figure 2), the idea being to refract each of the coloured rays a second time, in a direction at right angles to the former, and to see whether the blue rays, which were more refracted than the red, would be so again. Also to see whether they would retain their original colour, or whether they would undergo further analysis.

We will now repeat this experiment. I put in a second prism at right angles to the former, so that the light is thrown upwards. You see that the colours remain the same and in the same order as before. The blue remains blue and the red

remains red, but the blue is again more refracted than the red, and takes a higher position on the screen, the whole image or spectrum now taking an inclined position.

Newton made several more elaborate variants of this experiment. I will show you one of them. Two holes are used, with a prism in front of each, and two spectra are then projected on the screen in line, the red of one following on the blue of the other. Refracting them upwards with a third prism, we see that this line is dislocated, and that blue and red which were before adjacent are now widely separated (figure 3).
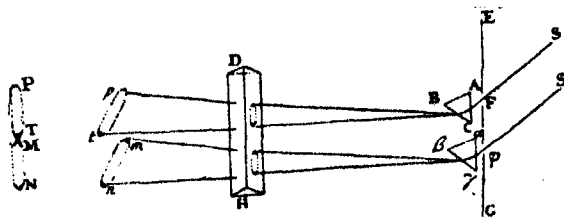
FIGURE 3

Newton in this way arrived at the view that the original white light of the sun consisted of coloured components, red, yellow, green, blue and violet. Each of these had a characteristic index of refraction and was bent to a perceptibly different extent by the prism. The result was that the original round white image of the sun was replaced by a series of coloured images side by side, overlapping and blending one with another. On introducing a second prism at right angles each of these coloured images was again refracted to its characteristic extent and each retained its colour without further analysis (figure 2).

Newton, however, soon saw the crudeness of his first arrangements, and we will now consider the methods he used to improve them. As he pointed out, the individual coloured images of the sun overlap at the middle, but not at the edges. Images of the sun are not in fact very suitable, because the angular width of the sun is about half a degree, and this is a considerable fraction of the angle between the red and violet rays. We want to make the sun in effect smaller, by blocking out part of its disk, which might be done in imagination by a distant diaphragm supported in mid-air. At the same time the hole in the shutter might with advantage be replaced by a lens, which would render the images sharper. When we have got so far in imagination, we notice that the long distance between the diaphragm in mid-air and the lens is not essential; with a suitable lens a moderate distance will do. We bring the diaphragm nearer, and fix it in the shutter, and bring the lens forward into the room, placing it so as to form a distinct picture of the diaphragm on the wall (figure 4).

We have now the diaphragm fixed in the shutter, and backed by the sun, as our effective source of light. In this respect it differs from the original diaphragm, which acted not only as a source, but as an image-forming pinhole. We now

have a picture of the diaphragm and not of the sun, and it will be advantageous to substitute a slit, which would not previously have been of any use, because the size of the image was then defined by the angular diameter of the sun. With the slit *and the lens* we shall get narrow well-defined images in each colour, and the colours will be pure and not overlapping. (Needless to say, modern refinements carry the matter much further.)

The explanation I have just given is in substance Newton's own.

We now adopt his improved arrangement, and you see the image of the slit projected on the screen. When I put in the prism, this narrow image is dispersed into a pure spectrum.
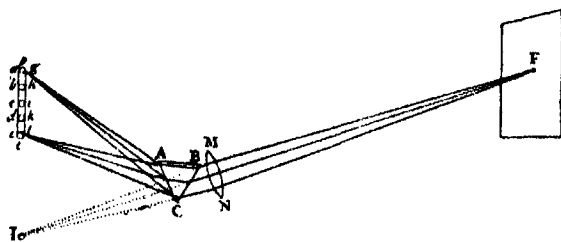
FIGURE 4
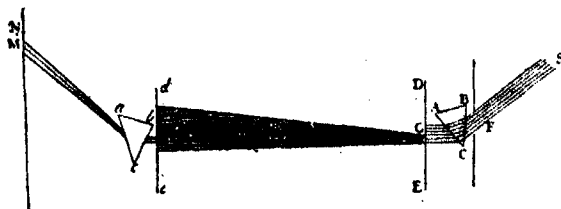
FIGURE 5

Let us now project the spectrum on to a small screen with a slit (figure 5) in it, which to avoid confusion I will call the screen-aperture. A colour passing through the screen-aperture is isolated, and you see it thrown on a second screen beyond. If I rotate the prism, I can throw the various colours in succession on the slit in the screen-aperture, so that the image on the second screen is taken through the succession of spectrum colours from red to violet. Starting with it at the red, I place a second prism in position behind the screen-aperture, and the rays are deviated, as we should expect. There is no elongation of the image this time, nor is there any diversity of colour. The red image remains red. The same applies to any other colour of the spectrum that is passed through the screen-aperture. But you will see that as the successive colours are passed, from red to violet, the deviation on the second screen continually increases, the red being the least deviated and the violet the most.

This experiment shows essentially the same thing as the experiment of crossed spectra which we saw before, but more perfectly.

The important point is that having once analysed white light into its prismatic constituent colours, this analysis is final, and further prismatic analyses can do no more. Further, each colour has its characteristic refrangibility, which will show itself at every refraction. Newton attached the greatest importance to this experiment, which he called, in Baconian phrase, the 'experimentum crucis'.

Having effected the analysis of white light into its constituent colours, he proceeded to reverse the process, and to show how the colours of the spectrum could be recompounded into white light. This he did by projecting the spectrum upon a large lens (figure 6) when the various rays were all converted to the same place, and produced a patch of white light at the focus. We will now repeat this experiment. If I place a white card immediately near the large lens, the spectrum is of course seen upon it. As we move the card away, the luminous patch contracts, and becomes white in the middle, though it remains red on the right and blue on the left. At the focus these coloured borders disappear, and there is nothing but white. Beyond the focus, where the red and blue rays have crossed, these colours are exchanged, and at longer distances still we have the spectrum inverted with the red on the left.
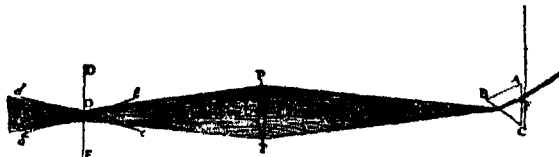
FIGURE 6
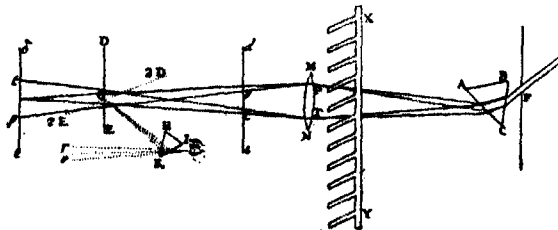
FIGURE 7

Let us now return to the focus where we have the colours converged to form white. I have here a kind of large-scale model of a comb as used by Newton (figure 7) which I can place in front of the lens. When I do so some part of the spectrum will be intercepted by the comb, and you see that the white patch produced by recombination of the entire spectrum loses its whiteness, and becomes tinged with colour. The colours produced in this way are not pure spectrum colours, because any tooth of the comb permits colour from parts of the spectrum on either side of it to pass. The colours therefore are composite, but we see that the complete spectrum is necessary to produce white. If the comb is moved slowly to and fro, you see that the coloured patch varies, being now reddish, now yellowish,

now bluish, but never white. If, however, it is moved to and fro as rapidly as may be, all the spectrum colours are present in succession. They blend by persistence of vision, and the patch becomes white again.

Newton devoted considerable attention to the colours of natural bodies, and he came to the conclusion that a coloured body is only coloured in virtue of its power of reflecting some of the components of white light more than others. I project the spectrum on a white screen. You note where the red light is seen. Now I substitute a red screen. We see red light where we saw it before, and nowhere else. The colours are not altered, though red is the only one which is strongly reflected by the red screen. The others are mostly absorbed, and their place in the spectrum is comparatively dark. Red is advantageous for such experiments, as red pigments are much the purest in colour.

In another of Newton's experiments the spectrum is projected on a large lens as before. If a white card is held at the focus, it appears white, for all the colours are converged upon it. If a card painted with cinnabar (vermilion) is placed there, it appears red as usual. If we block out the red part of the spectrum at the lens, the card shows no trace of red, becoming comparatively dark. If we block out everything but red, the card shows up with a purer red than at first.

We have now been through most of the experiments by which Newton supported his main proposition 'The light of the sun consists of rays differently refrangible,' following his own methods as closely as might be. They are a model for all time of how experimental research should be conducted, and it is difficult to our generation to see how any intelligent person could refuse his assent. This was by no means the unanimous opinion of his own contemporaries. One critic, Linus, by name, maintained that the spectrum could only be seen when the sun was shining through cloud, and it is strange, but true, that Newton was seriously discomposed by this ridiculous mare's nest. Goethe, as Professor Andrade has already mentioned, bitterly attacked Newton, using abusive expressions, and apparently thinking that to use isolated beams of sunlight admitted into a dark room was treating nature in an unfair manner, and compelling her under torture to say what she did not really mean!

The opposition to Newton's views did not finally die down till within measurable distance of our own time. In a book dated 1853, and dedicated to a man whom I as a boy personally knew, Sir David Brewster says: 'The conclusion deduced by Newton is no longer admissible as a general truth "That to the same degree of refrangibility ever belongs the same colour and to the same colour ever belongs the same degree of refrangibility".' *

Let us not make the mistake of thinking that Newton in his optical researches, exploring virgin ground, had easy triumphs. He did not escape the usual fate of discoverers. His jealous contemporary Hooke said his conclusions were not new. His admiring biographer, Brewster, concluded that they were not true!

* It is not intended to criticize Brewster adversely. The appearances by which he was misled are extremely deceptive.

Newton and the science of to-day

By Sir James Jeans, F.R.S.

We are met to commemorate the greatest of our men of science, and it falls to my lot to discuss where he stands in relation to the science of to-day.

Three hundred years have elapsed since he was born, and 276 since his fertile mind had conceived most of those fundamental ideas, the development of which was to add such lustre to his name and be of such outstanding value to science— in astronomy the idea of universal gravitation; in physics the theory of colours; in mathematics the differential and integral calculus (or, as he would have said, the direct and inverse fluxions) as well as the binomial theorem and the method of infinite series. He tells us he had thought of all these before he was twenty-four years old, although he wrote little about most of them until many years later.

The intervening years have seen Newton's ideas examined, tested and developed as no other set of scientific ideas ever has been, and it might seem that by now we ought to be able to assess his greatness and place him in his rightful position relative not only to the science of to-day, but also to that of all time.

We have no doubts as to his greatness, but we probably feel less confidence in our powers to assess his ultimate position in science than we should have done fifty years ago, and certainly less than his immediate successors felt. For they were not content merely to claim outstanding greatness for Newton; they went on to claim a quality of finality and uniqueness which we know better than claim for him to-day.

Laplace, for instance, wrote that the *Principia* was assured *for all time* of a pre-eminence above all other productions of the human intellect, while Lagrange, after expressing himself in similar terms, went on to say that Newton was 'not only the greatest genius that had ever existed, but was also the most fortunate, for, as there is only one universe, it can fall to only one man in the world's history to interpret its laws'.

Such eulogies as these are in striking contrast with Newton's own modest estimate of himself. As his life was drawing to its close, he made the much quoted remark that to himself he seemed only like a boy playing on the sea-shore, 'and diverting myself', he said, 'in now and then finding a smoother pebble or a prettier shell than ordinary, while the great ocean of truth lay all undiscovered before me'. Which of these two estimates, we may ask, is the nearer to the truth? And how far is it possible to reconcile them? Newton's contemporaries obviously could not have reconciled them; can we?

All work in mathematics, no matter how modest it may be, may claim immortality if only it is accurate, since absolute truth can have nothing to fear from the disintegrating influence of time. But it is different with work in astronomy

[251]

or physics. This must at best be relative to the knowledge, or even to the hypotheses, of its time, so that an increase of knowledge or an abandonment of hypotheses may at any time show it to be of no permanent value. This is as true of the work of Newton as of that of lesser men.

Yet Newton's contemporaries, as well as many generations after them, made the mistake of supposing that Newton's work in astronomy and physics possessed the same sort of absolute validity as his discoveries in mathematics; they thought that no future scientist would ever be able to penetrate to the innermost recesses of Nature's temple and discover more of her fundamental mysteries, because Newton had been there already and had seen all that there was to see. But we of to-day know that Nature's temple has many chambers, room after room, and that Newton had hardly penetrated further than the ante-chamber. Our thoughts turn to the work of Planck, Rutherford and Einstein. The ante-chamber which Newton had explored was familiar ground to them, but they found the keys to other, and hitherto unsuspected, rooms; in these they discovered laws every bit as fundamental as any that Newton discovered. We see that Newton was wiser than his contemporaries in his vision of the great ocean of truth which lay all undiscovered before him. But does this add to his intellectual stature or diminish it? Should it make him seem greater or less in our eyes?

There are some—although mostly laymen in science—who see science primarily as something that is for ever changing. For them the science of any period is like the sand-castles that the children build on the sea-shore; the rising tide will soon wash them away, and leave the sands clear for the new array of castles which will be built the next day. Those who hold such views are led, somewhat naturally, to make such statements as that Newton is out-of-date and superseded.

But the comparison is obviously faulty. Science is knowledge, and the primary characteristic of knowledge is not that it is for ever changing, but that it is for ever growing. Of course, all growth implies change, so that science must continually change—like the tree which is for ever budding out in new directions—but this is only a secondary effect. The correct comparison is not with sand-castles, which change because they are continually washed away and replaced, but with a vast building which changes as one floor is built on top of another, or a new wing is built where none stood before. This building is not like a medieval cathedral, the creation of a great number of artificers, each building after his own taste and fancy. It is an embodiment of scientific truth, and the truths of science are the same, no matter who discovers them. Whatever artificers build the structure, the blue-prints have previously been drawn by Nature herself, so that, whether Newton had lived or not, the building must at some time or other have looked pretty much as it looks to-day. And if we are to estimate how much modern science owes to Newton, we must not only consider how much Newton built himself, but also by how much he expedited the building—to what extent, to put it crudely, did Newton act as a catalyst to the growth and consolidation of scientific knowledge?

Let us first consider the significance of Newton's work in pure mathematics.

There have been instances in this subject—especially in the Theory of Numbers—where a worker has stated a theorem which he believed to be true, but which centuries of discussion have not succeeded in either proving or disproving. Newton's mathematical work was not of that type. His métier was not the discovery of fanciful but comparatively useless theorems in recondite branches of pure mathematics, but the forging of practical tools which were urgently needed for further scientific progress, and we may be sure that if Newton had not forged them someone else would have done so in time.

For instance, Newton hit upon the binomial theorem in two distinct ways. He, of course, knew the expansions of $(1-x)$ raised to integral powers—squares, cubes, etc.; these are easily found by simple multiplication. Starting from these known expansions, Newton interpolated to find the values of the coefficients when $1-x$ was raised to intermediate values, such as $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$. He also calculated the value of $(1-x)^{\frac{1}{2}}$ by taking the square root of $(1-x)$ by the simple rules of schoolboy arithmetic, and of course found that the result confirmed what he had already obtained. It is hard to imagine that such simple artifices could have remained untried for long; the binomial theorem was simply a nugget lying on the direct road of progress, and it is quite inconceivable that generations of mathematicians could have passed it by without noticing it.

It was much the same with the greater discovery of the method of fluxions. Rouse Ball tells us that, before Newton appeared on the scene at all, the method had been foreshadowed in the writings of no fewer than seven mathematicians—Napier, Kepler, Cavalieri, Pascal, Fermat, Wallis and Barrow. And in the more convenient and practical form of the differential calculus, it was actually discovered by Leibniz in 1675, and published to the world in 1684—nine years before Newton published his fluxional calculus. Apart from the personal questions involved, the bitter controversy as to whether the two discoveries were entirely independent, or whether Leibniz got ideas from seeing Newton's earlier letters on the subject, seems to me relatively unimportant; the calculus was simply another nugget lying on the direct line of scientific advance, and someone was bound to notice it and pick it up—if not to-day, then to-morrow; if not Newton, then Leibniz.

When we turn to consider Newton's still more famous work in mechanics and astronomy, we find the same story repeated; he did not drag science after him into peculiar paths of his own choosing, but kept to the broad main road, discussing the same problems as others were already discussing. The fundamental ideas with which he worked were not in any way novel, or peculiar to himself. His first law of motion had been given by Descartes in 1644, and indeed Plutarch had given it 1500 years earlier together with a foreshadowing of the general principles of planetary motions, and the conception of gravity extending as far as the orbit of the moon. In his *De facie in orbe lunae*, Plutarch had written, 'You fear that the moon may fall, on the grounds that only light air circulates under the moon and that this is not adequate to bear a solid weight. But the moon is

secured against falling by her motion and the swing of her revolution—just as objects put in slings are prevented from falling by the circular whirl. For everything is carried along by the motion natural to it, if it is not deflected by anything else.'

Much the same, again, may be said about Newton's famous law of gravitation. The attraction of lodestone for iron had familiarized men with the idea that matter could attract other matter across empty space and, as far back as 1600, Gilbert in his *de Magnete* had conjectured that the earth might attract the moon like a magnet. Kepler had gone further in the same direction in 1619; we find him insisting on the mutual attraction of all matter, instancing the tendency of all objects to fall towards the centre of the earth, as well as the ocean tides which, he maintained, were caused by the attraction of the moon. He said that two stones out in free space would approach one another, like mutually attracting magnets, and would finally meet at the point which we now call the centre of gravity of the two. He thought that the planets were kept moving in their orbits by some power inherent in the sun, and, although forming a very wrong idea of the nature of this power, he conjectured that it must vary as the inverse square of the distance. However, he immediately abandoned this law in favour of one varying inversely as the simple distance, a change which was challenged by Bouillard, who insisted that the true law must be that of the inverse square. Indeed, one of Kepler's own laws—that connecting the periodic times of the planets with their distance from the sun—shows clearly that the true law cannot be other than the inverse square law. Newton tells us that he had already noticed this in the plague years 1665 and 1666, but he did not publish it until it had been noticed independently by Hooke, Halley, Huygens and Wren.

Looking back on all this, we can hardly accept Wordsworth's description of Newton's mind as

'for ever voyaging through strange seas of thought alone'.

Except in optics, where he voyaged alone because everyone else had lost the way, Newton's ideas were very much those of his contemporaries. His main achievement was not, as is so often stated, that he was the first to think of gravity as extending to the orbit of the moon, for this had been thought of 1500 years before he was born. Neither was it that he was the first to think of universal gravitation, for many had thought of it before him. Neither was it that he was the first to conjecture that the force of gravitation must fall off as the inverse square of the distance, for this idea was generally current in his time. It was not even that he was the first to understand the general principle of planetary motions, for others understood these equally well.

His great service to science, and through science to the human race, was not that he for ever voyaged through strange seas of thought alone, but that he voyaged through familiar seas of thought—if not in company with others, at least in seas that were much frequented by others. Bacon had written of his own

contributions to scientific thought, that they did not originate in any mental powers of his own, but rather in the spirit of his age—'partus temporis potius quam ingenii'. It was this same spirit of the age that moulded the form of most of Newton's contributions to science. Scattered scientific ideas—and unproved conjectures in particular—were floating around in abundance, and lesser lights of the firmament had made small, hesitating and uncertain advances in many directions. The age was crying for a man who could systematize, synthesize and extend the whole, and it found him in superlative excellence in Newton. But beyond all this, mental powers such as Bacon had disclaimed for himself were present in overflowing measure in Newton—'qui genus humanum ingenio superavit'.

His special talents, as I see them, were: first that, out of a large mass of confused ideas provided largely by others, he was able, with his clear and acute mind, unerringly to pick out the true and discard the false; second that, having done this, his amazing mathematical ability enabled him to replace conjecture by proof, and so provide a firm basis from which he could unhesitatingly advance; third that he did advance, with quite incredible speed, sureness and directness, his supreme mathematical ability enabling him to outstrip all competitors with the utmost ease. From *conjecturing* that a thing is so to *knowing* that it is so, is usually a very long journey; it is one which Newton was so often able to take when his competitors were not. For instance, Kepler had *conjectured* that the law of gravitational force might be the law of the inverse square; Newton *knew* that it could not be anything else. Kepler was so far from certain knowledge that he immediately abandoned this conjecture for another; Newton could not be misled into so doing, because his mathematical investigations provided him with convincing proof of the truth of his ideas. And having once settled his fundamental ideas and feeling full confidence in them, he was able to stride forward into the unknown with sure and rapid steps until he had unravelled almost the whole mystery of the then known universe.

In general it was not in originating ideas but in developing them that his greatness showed itself most outstandingly; he was primarily a synthesist and systematizer. Yet here, as ever when we try to compress his genius into a formula or to map out its boundaries, an exception turns up to suggest that it overflowed all limitations; here it is the optics.

If we wish to understand the fundamental cause of his success, we cannot do better than read what Macaulay wrote in his *History*: 'In Isaac Newton two kinds of intellectual power which have little in common, and which are not often found together in a high degree of vigour, but which are nevertheless equally necessary in the most sublime departments of physics, were united as they have never been united before or since. In no other mind have the demonstrative faculty and the inductive faculty coexisted in such supreme excellence and perfect harmony.' Einstein delivers a similar judgement even more forcibly: 'In one person he combined the experimenter, the theorist, the mechanic and, not least, the artist in exposition. He stands before us, strong, certain and alone: his joy in creation and

his minute precision are evident in every word and in every figure.' And in truth
he was unmatched both in the breadth and depth of his powers. Both contributed
to his success, but if anything, the breadth was more in requisition, and so made
the greater contribution to the success.

Yet if we are to estimate the extent to which Newton, with his outstanding
mental power and unique combination of talents, expedited the growth of scientific
knowledge, we cannot but place to the other side of the account his unfortunate
habit of keeping his results to himself until they had been discovered independently
by someone else. To some extent, this also may be ascribed to the spirit of the
age. The seventeenth century had but little of our vision of science as a great
benefactor to the human race, or of new knowledge as something which ought to
be disclosed at once so as to expedite further progress by others. A man's scientific
discoveries were his private property, with which he could do as he liked. Often
his ambition was not to enable others to benefit from them, but to prevent others
benefiting from them—this is why he would often publish his results in cypher.
Or he might regard science, and mathematics in particular, as a sort of mental
gymnasium in which contests of skill took place, and competitors issued challenges
to one another to determine which of the two was the stronger or quicker. To use
Newton's own metaphor, science was, more than to-day, a matter of playing with
pebbles on the seashore, and seeing who could find the smoothest pebble or the
prettiest shell. Thus there was less feeling of moral responsibility than now, perhaps
even less interest in science for its own sake.

And so we see Newton's terrifically powerful mind playing with the problems
of science like a cat which plays with a mouse and loses all interest when he has
killed it, or as we play with a crossword puzzle and regard the incident as finished
when we have solved it.

We can see this exemplified in nearly all Newton's major discoveries. He tells
us that he had thought of the method of fluxions in the plague year of 1665. Yet
before publishing it even partially (except in cypher) to the world, he allowed
28 years to elapse—years in which Leibniz published and discovered the same
thing—and it was not fully published until 1736. Again in this same year of 1665,
Newton satisfied himself that a force of gravity obeying an inverse square law
explained the motion of the moon 'pretty nearly', and was content to leave it at
that. Later, he found that this same law would result in the planets describing
the exact orbits they were known actually to describe—ellipses with the sun in
one focus. Again he was content to leave it there. Most men would have published
this as proof positive of the law of the inverse square, but Newton kept it to
himself until Halley travelled to Cambridge to consult him on this very problem.
Newton then explained that he had solved the problem five years previously, but
had mislaid his proof. Finally Professor Andrade has reminded us of the reluctance
with which Newton published his *Principia*—probably the greatest single work of
the human intellect—and of the pressure which had to be applied before he
would do so.

Apart from the general secretiveness of his age, we may perhaps find three further causes for this delaying tendency in Newton: an overactive mind which refused to halt by the wayside to perfect and write out results, a disinclination to be satisfied with anything short of perfection, and an extreme sensitiveness to criticism and consequent desire to avoid controversy, this resulting more than once in a disinclination to pursue the study of science any further. In spite of these handicaps Newton bestrode the scientific world like a Colossus: what would he have been without them?

Such was the man as I see him; let us now consider how much of the scientific edifice which Newton built, or was built under his influence, is still in useful service to-day.

Outside pure mathematics, Newton's procedure was simply that of every other true man of science. He started from known facts which, in his judgement, seemed likely to be related—as for instance, if legend can be trusted, that apples fall to the ground with an acceleration of 981 c.g.s. units, and that the moon circles round the earth once every 27⅓ days—and tried to represent the isolated facts of such a group as special instances of a general system of laws. If this could be done, the set of laws so discovered could be made to predict other phenomena, which could then be submitted to the test of experiment or observation. If the laws passed this test, they were assumed to correspond to something in ultimate reality; provisionally at least they were true laws, and showed how nature worked. If Newton differed from others, it was in the greater range of facts on which he worked.

In any scientific era, the search for such sets of laws is clearly restricted by the range of facts which are known, or are capable of being known, at the time. And we cannot expect scientists to discover laws of a higher order of accuracy than the observations through which the laws are discovered.

Some may challenge this, arguing that a genius of sufficient stature ought to be able to discover the true reality behind nature from quite meagre evidence—'ex pede Herculem'—and that, when this reality was known, the laws governing its behaviour could be deduced and would necessarily be true throughout the whole range of phenomena, whether known or still to be discovered. Unhappily nature is not like that. Science can, from the nature of things, never attain to a knowledge of the realities of the world, but only to a knowledge of the phenomena—of the impressions that the world makes on our senses, generally of course through instruments of precision. Our sense-organs form a sort of screen on which nature's lantern is for ever projecting pictures; we can study these pictures, but can never pass behind the screen to see how the lantern works. And as the sequence of pictures thrown on the screen is all that we can ever know or study, a set of laws that links the pictures together in a perfect systematic order is the most that science can provide, or can properly be asked to provide. If meagreness of the instrumental equipment in any age prevents the science of that age from knowing the finer details of the pictures, then we cannot expect the science of the age to provide laws governing these fine details.

For this reason we have to introduce a sort of relativity into our conception of the aims of physics; we may say that the science of any age can aim no higher than at predicting the result of any experiment that can be performed *in the age in question*. If this is conceded, we may claim that by far the greater part of Newton's work was accurate and final, relative to the age in which he lived. And this part is still of value and still in use to-day.

The enormous increase in instrumental power which has occurred since Newton's time has thrown open new fields for exploration in two directions—towards the infinitely great and towards the infinitely small, from the electron at one end of the scale to the nebulae and the whole of space at the other. Man and the man-sized world with which Newton was mainly concerned lie about midway between these two extremes.

The laws of nature are of course universal, so that the same set of laws must prevail throughout the whole of this range, but different aspects of these laws assume importance in turn in different parts of the range. So much is this the case that we may almost regard the different parts of the range as constituting separate and detached worlds in which completely different sets of laws prevail. There is the small-scale world of electrons and of atomic physics in general in which the laws of quantum mechanics prevail, the man-sized world in which the laws of molar mechanics prevail, and the world of the great nebulae in which the laws of relativity prevail. These three worlds are all governed by the same laws, but factors which are all-important in one become mere insignificant corrections in the others. Newton's work was applicable almost exclusively to the middle-sized world of molar mechanics, and so not in general to the other two worlds, the existence of which was hardly suspected in his day. Lagrange was wrong in thinking that there was only one universe, and that it could fall to only one man to interpret its laws; there are worlds within worlds, and Newton had only interpreted the laws of one of these. Relative to this world most of Newton's work was accurate and final, which means that it was also accurate and final relative to the age in which he lived.

If we examine any recent paper on theoretical physics, we are likely to find it plentifully sprinkled with either the symbol h or the symbol c, or both. These two symbols were unknown to Newton, at least in the sense in which we now use them; they are, so to say, the emblems of the theories of quanta and relativity which did not invade physics until the present century.

The symbol c denotes the velocity of light, which is greater than the velocity of any material object, and is usually enormously greater. Also, in dynamical problems the ratio of these two velocities can only enter through its square. We are now familiar with electrons which move almost as fast as light, but the kind of dynamics with which Newton was concerned contemplated no velocity greater than that of the planet Mercury, which is about 48 km. a second. For this, the ratio of the squares of the two velocities is 0·0000000256, or, to seven significant figures, is *nil*. If we disregard this fraction, the dynamics of relativity becomes

absolutely identical with that of Newton, so that to seven significant figures Newton's dynamics was adequate and factually perfect for an age in which nothing moved faster than the planet Mercury.

It is the same with the h of the theory of quanta. This measures the atom of action which Planck discovered in 1899, and although it is of the utmost importance to phenomena on the atomic and subatomic scale, it is absurdly small in comparison with the amounts of action involved in the activities of everyday life. Roughly speaking, the ratio is that of an atom to a gramme, or of a gramme to a star. This small quantity again was quite negligible in all the problems with which Newton was concerned.

Thus the Newtonian mechanics, dynamics and astronomy were absolutely right factually, except for quantities which were so small that they could not either affect Newton's problems in any way, or make their existence known through the best instrumental accuracy that was available in his day. To see Newton's work in these subjects in their proper perspective, we must think of him as the builder of the ground floor of the great building of mechanics, dynamics and astronomy, and as its architect for many floors above this. Newton's first and last serious contribution was his *Principia*, but after this floor after floor was added on strictly Newtonian lines—on the dynamical side, the generalized dynamics of Lagrange, Hamilton and Jacobi; on the astronomical side, the dynamical astronomy of Laplace and many others; on the physical side, the electrodynamical theory which we associate primarily with Clerk Maxwell. All this and much more followed the Newtonian architectural plan without any deviation whatever for two centuries. It was not until 1887—just 200 years after the appearance of the *Principia*—that the Michelson-Morley experiment first showed that the Newtonian scheme was in actual fact imperfect. But the experiment had to be discussed for eighteen years before this inference was drawn, and it was only when Einstein's restricted theory of relativity appeared that the Newtonian scheme was generally recognized to be inadequate.

It is, however, inadequate only with reference to the ultra-refinements of modern science. When the astronomer wishes to prepare his *Nautical Almanac*, or to discuss the motions of the planets, he uses the Newtonian scheme almost exclusively. And the engineer who is building a bridge or a ship or a locomotive does precisely what he would have done had Newton's scheme never been proved inadequate. The same is true of the electrical engineer, whether he is mending a telephone or designing a power-station. The science of everyday life is still wholly Newtonian. And it is impossible to estimate how much this science owes to Newton's clear and penetrating mind having set it on the right road, and this so firmly and convincingly that none who understood his methods could doubt their rightness. Without his guidance, smaller men might have argued for centuries as to what was right and what was wrong. When Newton had spoken, it soon came to be recognized that the time for controversy as to which was the right road was past; it only remained to advance along the Newtonian road.

Newton's work in optics stands somewhat apart from the rest in that he was much less indebted to his predecessors and contemporaries. Indeed, current ideas on the nature of light were so vague and erroneous that he was probably hindered rather than helped by what he knew of them, and was perhaps fortunate in not knowing more. We know that he attended the lectures of Barrow in Cambridge, where he would doubtless become acquainted with the optical theories of Descartes as well as with those of Barrow himself. Now Descartes thought that all space was filled with substance, and that vision resulted from the transmission of a pressure from one particle to another of this substance—much as a blind man pokes about with a stick. He thought that different sensations of colour resulted from the particles rotating with different speeds, red resulting from the most rapid rotations, and yellow, green and blue following in this order. Barrow held even more fantastic opinions; he thought that red light was more concentrated than light of other colours, but was broken up by intervals of complete darkness, yellow light consisted of a mixture of this red light with white, and so on. Hooke conjectured that light consists of a rapid vibrational motion of an ethereal medium filling all space, different colours being produced by different shapes of wave; blue, for instance, resulted from 'an oblique and confused pulse of light'.

Into the midst of all this confusion of thought came Newton, to whom, in Einstein's words, 'Nature was an open book, whose letters he could read without effort. The conceptions which he used to reduce the material of experience to order seemed to flow spontaneously from experience itself, from the beautiful experiments which he arranged in order like playthings.' With the prism he had bought at Stourbridge Fair, he went at once to the crucial experiment of the spectrum. Finding that this was longer than it was broad, he at once saw that differences of colour must result from different degrees of refrangibility. This discovery alone, the subject of Newton's first scientific paper, published in our *Philosophical Transactions* of 1672, at once took the theory of light leagues beyond the point to which any previous investigator had brought it, and opened the road to further rapid progress.

But Newton brought his views into discredit in the eyes of many by appearing to think of light as a material substance—Hooke definitely brought this charge against him. Actually Newton repeatedly claimed that his results did not depend on any special views as to the ultimate nature of light, and he was careful to avoid such words as 'corpuscles' which might seem to imply such views. Yet much of his thought seems to become meaningless unless we identify his rays—the 'least parts' of light—with something of a corpuscular nature. However little he may have said about it, Newton seems always to have had his own picture of the structure of light before his eyes, and this picture was corpuscular.

For two centuries many thought of this as the one serious mistake that Newton ever made. Recently the pendulum has swung—perhaps too far—in the other direction. Light is a transfer of energy from matter at one place to matter at another, and this transfer can only take place by complete units, or 'quanta'.

This led Einstein to picture light as a shower of arrows, each arrow containing a single complete quantum of energy, and some have hailed Newton's corpuscular theory of light as a brilliant prevision of this.

But we must remember that Newton introduced his corpuscles primarily to explain the rectilinear propagation of light, whereas Einstein introduced his light-arrows to picture the quite different property of atomicity. This being so, the similarity of the two pictures can hardly be quoted as a supreme example of intuition on the part of Newton. It might be if light really did consist of corpuscles; Newton would then have arrived at the truth by one road 200 years before Einstein reached it by another. But no competent physicist can for a moment suppose that light consists either of corpuscles or of light-arrows; these are only pictures we draw for ourselves to help us understand parts of a truth which we know, as a whole, to be for ever beyond our comprehension.

At this point we are inevitably brought to consider Newton as a philosopher. Unlike his great predecessor Descartes and his great contemporary Leibniz, he seems to have had very little interest in philosophy for its own sake. He was what we might describe to-day as a pragmatist; his main concern with a philosophy was not as to whether it was true, but whether it worked and yielded useful results. It was another illustration of the ever-recurring difference between the British and the Continentals.

Descartes had taught mathematicians to measure co-ordinates from fixed lines on a piece of paper. Newton wanted to be able to measure mechanical and astronomical co-ordinates in the same way from positions in space and instants in time. So he assumed without further ado that time could be measured from a fixed standard instant, and that the remotest parts of the universe contained vast immovable masses, providing standard positions against which absolute motion could be measured.

This assumption of the existence of absolute time and space was a mere working hypothesis which, as he admitted, might not be true at all: 'it may be that there is no body really at rest, to which the places and motions of other bodies can be referred'. But the assumption, whether true or false, produced results—nothing less than a consistent scheme which explained a large part of the then known workings of nature. This being so, philosophers such as Bruno and Leibniz might argue in vain that absolute space, time and motion were meaningless and illogical.

It was much the same in mechanics. In the man-sized world of seventeenth-century science, motion was the result of pushes or pulls of the kind that a horse exerts on a cart. Other suggested causes of motion, such as action at a distance, were regarded much as we of to-day regard levitation or table-turning, and described as 'occult qualities'. It was inevitable that Newton's force of gravity should come under this description. Leibniz wrote 'Some men begin to revive, under the specious name of forces, the occult qualities of scholasticism, but they bring us back again into the Kingdom of Darkness'. Newton's reply was a fine defence of his pragmatical philosophy. To understand the motions of the planets under

gravity, he says, even though we do not know the cause of gravity, is as good progress in philosophy as is a knowledge of the motions of the wheels of a clock in the philosophy of clockwork, even though we do not understand why the weight which moves the wheels falls earthward.

At the same time, he refused to admit that his gravitational forces were not wholly mechanical. From the celestial phenomena, he said, we calculate the force of gravity with which bodies are drawn to the sun and the several planets. Once having found these, we deduce the motions of the planets, the comets, the moon and the sun, and 'I wish we could derive the rest of the phenomena of nature by the same kind of reasoning from mechanical principles'.

It is interesting to speculate as to what might have happened if Newton had shared the views of Leibniz as to the relativity of all motion. It is conceivable that he might have arrived speedily at what we now call the restricted theory of relativity, that through it he would have seen the impossibility of action at a distance, and would then have developed a relativity theory of gravitation similar to that of Einstein. It is conceivable, but I can hardly think it would have happened—the necessary physical ideas and mathematical technique seem to be centuries removed from the fields in which Newton worked. It is more likely that he would have been compelled to come down to an approximation which would have brought him back to the Newtonian mechanics and the inverse square law. If so, his erroneous philosophical background of absolute space and time was a real advantage, since it provided him with a short cut to a region which would have had to be explored anyhow as a preliminary to further progress. It enabled him to go direct to the main point as it appeared in the seventeenth century, neither hindered by any philosophical subtleties nor haunted by any fears that his system might not be an expression of final absolute truth.

In this way Newton created a system which, although philosophically unsound, was destined to endure for two centuries before any chinks were found in its armour. Then more philosophically minded Continentals had to be called in to patch up the deficiencies.

In this quality of practical and opportunist directness, as in many others, Newton seems to me to be typically British. In general he was strong where the British are strong, and weak where the British are weak. If we think of his various pre-eminences in turn, and consider who runs him closest in each, we shall find it is usually a British scientist who comes to mind. If we try to match the simple directness and the clear logical sequence of his experiments in optics, our thoughts turn to our own Faraday and Rutherford. Clerk Maxwell runs second to him, I think, in other aspects of his work. But if we think of those who excelled him in his philosophical outlook, we think exclusively of Continentals, from Leibniz on. His successes were typically British, and his failures—in so far as he ever failed—were also typically British. We can not only be proud that this greatest of all scientists was our own countryman, but we may reflect with satisfaction that he was no freak genius or sport, but was the concentrated embodiment of all the distinguishing characteristics of British science.

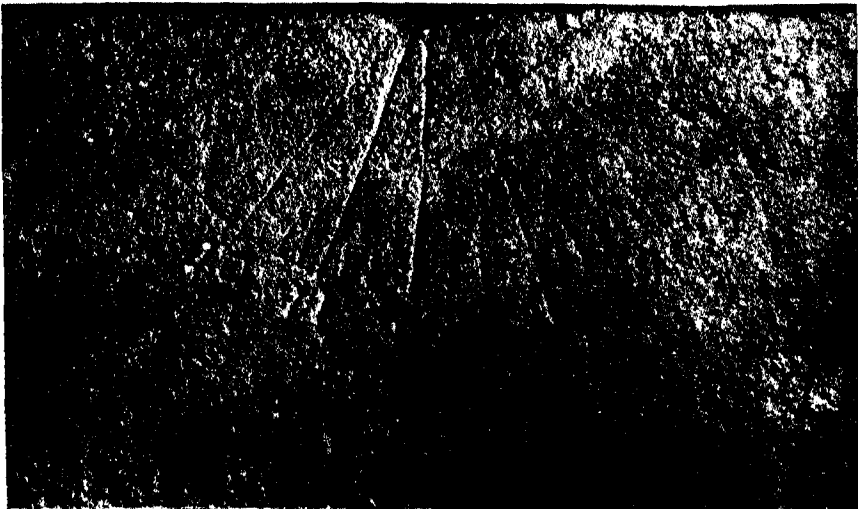ISAAC NEWTON
from a portrait by Johan Vanderbank in the possession of the Royal Society,
presented by Martin Folkes, P.R.S.

REFLECTING TELESCOPE
made by Isaac Newton with his own hands, 1671. Presented to the
Royal Society by Messrs Heath and Wing, 1766.

SOLAR DIAL
made by Isaac Newton when a boy, taken from the wall of Woolsthorpe Manor
House. Presented to the Royal Society by C. Turnor, F.R.S.

Exhibits

EXHIBITS ILLUSTRATING SOME OF NEWTON'S EXPERIMENTS

ARRANGED BY LORD RAYLEIGH

as a supplement to those shown in his Lecture on

Newton as an Experimenter

EXHIBIT No. I. *Opticks.* Book I. Experiment I. To show that lights which differ in colour differ also in refrangibility.

Newton's account abbreviated:

I took a black oblong stiff paper. One part I painted into a red Colour, the other into a blew. This paper I viewed through a Prism of Solid Glass. I found that if the refracting Angle of the Prism be turned upwards so that the paper may seem to be lifted upwards by the Refraction, its blew half will be lifted higher by the Refraction than its red half. But if the refracting Angle of the Prism be turned downwards, so that the Paper may seem to be carried lower by the Refraction, its blew half will be carried something lower than its red half.

EXHIBIT No. II. *A* and *B*. *A. Opticks.* Book I. Part II. Experiment XVI. In this experiment a uniform field of light is viewed through a reflecting prism. Newton used the cloudy sky. An illuminated opal glass is here substituted. The bright part of the field is the region of total reflexion. The relatively dark part of the field, corresponding to steeper angles of incidence, is the region of partial transmission and partial reflexion. Observe the blue arc marking the limit of total reflexion. This shows that white light contains a constituent (blue) which begins to be totally reflected at angles of incidence less steep than are necessary for the other components.

This blue colour being made by nothing else than by reflexion of a specular superficies seems so odd a Phenomenon, and so unaccountable for by the vulgar Hypothesis of Philosophers that I could not but think it deserved to be taken notice of.

B. A supplement is arranged to this exhibit, allowing the transmitted light to be examined. A pair of right-angled prisms is used forming a cube after the manner of Newton, reflexion occurring at the diagonal interface. The field is divided into a bright and a dark part as before, the dark part (no transmission) being the area of total reflexion. A reddish yellow arc marks the limit of transmission, which is complementary to the blue arc seen by reflexion.

This experiment was performed by Newton in a more elaborate way, using the sun's light and projecting a spectrum of the beam transmitted by the cube. The present arrangement is adopted to avoid the complication of an arc lamp, and reproduces Newton's essential point.

[**263**]

EXHIBIT No. III. *Opticks*. Book II. Part I. Observation xiii. This experiment shows that Newton's rings, formed between a spherical surface of large radius and a flat plate, are larger in red light than in blue. By pulling a string the observer may rotate the spectrum-forming prism and observe the rings dilating as the colour is changed from blue or blue green towards red. On releasing the string the colour passes back towards blue and the rings contract again.

Appointing an assistant to move the Prism to and fro about its Axis, that all the Colours might successively fall on that part of the Paper which I saw by reflexion from that part of the Glasses where the Circles appeared, so that all the Colours might be successively reflected from the Circles to my Eye whilst I held it immoveable, I found the Circles which the red light made to be manifestly bigger than those which were made by the blue and violet. And it was very pleasant to see them gradually swell or contract according as the Colour of the Light was changed.

Newton recognized that this proved that there was a coarser structure associated with red light than with blue. In Query 13 at the end he says:

Do not several sorts of rays make vibrations of several bignesses, which according to their bignesses excite sensations of several colours?

EXHIBIT No. IV. *A, B* and *C*. This is an experiment of Newton's on frictional electrification. (Description below.) Rub the upper surface of the glass with flannel, or in the case of *C* stroke it with the dry hand; and the movements of the pith or paper fragments will be observed. *C* is the easiest and should be tried first.

A. An apparatus with flint glass as used by Newton. The action of this is uncertain, depending on the hygroscopic condition of the glass. It may fail, as it did when first tried at the Royal Society. Application had to be made to Newton for further directions. In dry weather and with vigorous rubbing it works admirably. Try reversing the glass and rubbing first one side, then the other.

B. An apparatus with silica glass. This can be taken out and warmed over a spirit lamp. It will then work without fail.

C. An apparatus with a sheet of celluloid replacing the glass. This will work at any time without fail, if lightly stroked with the dry hand.

From the Minutes of the Royal Society. 9 December 1675.

That [Newton] having laid upon a table a round piece of glass about two inches broad, in a brass ring, so that the glass might be about one-third of an inch from the table, and the air between them inclosed upon all sides after the manner as if he had whelved a little sieve upon the table: and then rubbing the glass briskly, till some little fragments of paper, laid on the table under the glass, began to be attracted and move nimbly to and fro; after he had done rubbing the glass, the papers would continue a pretty while in various motions; sometimes leaping up to the glass and resting there awhile; then leaping down and resting there, and then leaping up and down again; and this sometimes in lines perpendicular to the table sometimes in oblique ones; sometimes also leaping up in one arch and down in another divers times together, without sensible resting between; sometimes skip in a bow from one part of the glass to another, without touching the table; and sometimes hang by a corner and turn often about very nimbly, as if they had been carried about in the midst of a whirlwind; and he otherwise variously moved every paper with a diverse motion. And upon sliding his finger on the upper side of the glass, though neither the glass nor inclosed air below were moved thereby, yet would

the papers, as they hung under the glass receive some new motion inclining this or that way, according as he moved his finger.

The experiment he proposes to be varied with a larger glass placed farther from the table, and to make use of bits of leaf gold instead of papers, esteeming that this will succeed much better, so as perhaps to make the gold rise and fall in spiral lines or whirl for a time in the air, without touching the table or glass.

Ordered that this experiment be tried the next meeting.

EXHIBITS ARRANGED BY THE ROYAL SOCIETY

1. The manuscript of the *PRINCIPIA*, from which the first edition was printed, with corrections by Newton.

2. The *REFLECTING TELESCOPE* made by Newton with his own hands in 1671.

3. A *SOLAR SUNDIAL* made by Newton when a boy. It was formerly in the wall of his home, the Manor House, Woolsthorpe.

4. A testimonial given by Newton in favour of William Jones (later F.R.S.), an applicant for a post on the staff of Christ's Hospital.

5. An *ASSAY FURNACE* used by Newton in His Majesty's Mint in the Tower of London for his experiments in cupellation.

6. A plan of His Majesty's Mint in the Tower of London, drawn for Newton on his appointment as Master Worker.* It is dated January 1701.

7. A holograph draft of Newton's report on *The value of gold in proportion to silver in several parts of Europe.*

8. Nine letters from Halley to Newton relating to the publication of the first edition of the *Principia.* They were written between 22 May 1686 and 5 July 1687.

9. *The hunting of ye green lyon* and *The standing of ye glass for ye time of putrefaction & congelation of ye medicine.* Alchemical verses copied by Newton with his own hand.

10. An *IVORY BUST* by Cheverton from the sculpture by Roubiliac.

11. *THE MANOR HOUSE OF WOOLSTHORPE.* A drawing in *Memoirs of Sir Isaac Newton's Life*, written by William Stukeley in 1752.

12. The mask of Newton's face, which belonged to Roubiliac, from the cast taken after death.

---

* The Deputy Master of the Mint has supplied the following note: 'In the 18th century the Royal Mint operated, in form, under the constitution laid down by Edward I. The Master Worker was a salaried officer of the Crown who contracted for the execution of the coinage, in law for the whole, in practice for a large part of the work. The Warden was a check on some of his actions, the Comptroller on others. Sir Isaac Newton was Warden from March 1696 to December 1699 and Master Worker from December 1699 to December 1726.'

ACKNOWLEDGMENTS

The cascade theory with collision loss

By H. J. Bhabha, F.R.S. and S. K. Chakrabarty

Electrons are assumed to suffer a constant energy loss β by collision, and the radiation loss and pair creation are taken to be described by the formulae of Bethe and Heitler valid for complete screening. With these assumptions a solution of the cascade equations is given in the form of a series, and it is shown that the series is so rapidly convergent that in general it is necessary only to calculate the first term. Collision loss enters into each of the terms in an essential way, and as a result the first term alone gives to a very considerable degree of accuracy *the whole energy spectrum of electrons from the highest energy to energies far below the critical energy.* For thicknesses greater than 1·5 in the characteristic unit of length the number of particles of energy E increases monotonically with decreasing E, but the spectrum gets flattened for energies below the critical energy. For thicknesses t below 1·5, the spectrum has a very different shape, decreasing first as E decreases from the primary energy and then increasing again to the smallest E, the flattening taking place now only for $E < \beta t$. It is shown that neglect of collision loss sometimes causes the number of electrons of even the critical energy to be as much as seven times too large. Tables of the spectra of cascade electrons due to primaries of different energies are given for five typical thicknesses.

The solution is also valid when the energy of the primary electron starting the cascade is comparable with or lower than the critical energy, and gives in a compact form the complete solution of the problem of the absorption of a low-energy electron by collision loss and cascade production.

It is generally accepted that the cascade theory put forward by Carlson & Oppenheimer (1937) and Bhabha & Heitler (1937) correctly describes all the general features of the behaviour of the soft component of cosmic radiation. Moreover, it appears to be in rough quantitative agreement with the observed absorption curve of cosmic radiation in the atmosphere and the production of electron showers in heavy materials by fast electrons. Nevertheless, a very exact comparison between theory and experiment has not yet been made, due, on the one hand, to inaccuracies in the theory introduced by the inexactness of the physical assumptions and approximations in the mathematical treatment, and on the other hand to uncertainties in the experimental data itself. A stage is, however, being reached when the gradually increasing data on the atmospheric absorption curves at different latitudes makes a more accurate knowledge of the theoretical behaviour of cascade processes necessary in order that a more exact comparison between theory and experiment may be made with a view to finding out whether experiment reveals other processes besides those explained by the cascade theory, and to what extent. The purpose of this paper is to give a more accurate theoretical treatment of cascade processes than has been done hitherto.

All the previous theoretical treatments of the subject have been inexact, first, because approximations have been made in describing the physical processes which play a role in the cascade theory, and secondly, because the mathematical treatment has sometimes not been exact even on the basis of these physical assumptions. The weakest feature in all the previous treatments has been their very inadequate

consideration of the effect of the ionization or collision loss suffered by the cascade electrons. In the paper of Bhabha & Heitler the concept of critical energy was introduced, defined as the energy of an electron for which its radiation loss is equal to its collision loss in the substance, and it was assumed that the collision loss would not appreciably affect the cascade process above the critical energy, while it would materially reduce the multiplication of electrons below the critical energy and lead to an absorption of the cascade. But the collision loss was not treated quantitatively. An attempt to calculate the number of electrons below the critical energy more accurately was made by Arley (1938) on the basis of very rough physical assumptions. The more accurate calculations of this paper show, however, that the figures he got for the distribution of electrons below the critical energy are not even qualitatively right for small thicknesses,* and give a justification for the very rough assumption of Bhabha & Heitler that for all except extremely small thicknesses the number of particles per unit energy range below the critical energy always increases slowly with decreasing energy.

Carlson & Oppenheimer, and later Snyder (1938) and Serber (1938) have given approximate formal solutions of the cascade equations with collision loss in the form of double integrals in the complex plane. These expressions cannot be calculated directly, and for evaluating the number of cascade electrons of energy E, they all get in essence an infinite series in powers of β/E, where β is approximately the critical energy. Closer analysis shows that these series are *divergent for all values of E*, and the one which is based on the most accurate physical assumptions, namely, Serber's series, is essentially the formal solution given by equation (42) of this paper, which we have rejected for the same reason. It will appear that this series cannot even be considered as an asymptotic solution of the problem when $E \lesssim \beta$. It can be shown that the divergence of Snyder and Serber's series is intimately connected with the fact that their solutions do not satisfy the correct boundary conditions at the surface of the layer. In our opinion the divergence of their series makes it impossible to say that they satisfy the actual boundary conditions even approximately. Snyder and Serber have used their solutions for calculating the total number of particles of all energies at any depth, but the serious defects mentioned above seem to us to throw doubt on the reliability of their numerical results. It is also obvious that even without the above defects all these methods would completely fail to give the energy spectrum of electrons with energies in the neighbourhood of the critical energy, which is a region of very great interest both theoretically and from the experimental point of view.

It is clear that a solution of the cascade problem valid over the whole range of energies cannot be obtained in the form of a series in powers of β. For if the collision loss be neglected, then the spectrum of electrons is of the form $1/E$ for small E and

* Arley has completely neglected the production of electrons below the critical energy by pair creation. This process is just the one that gives the most important contribution at small thicknesses, as will be seen in the last section of this paper. But his treatment of radiation processes is also much too crude to claim any quantitative accuracy.

hence tends to infinity as $E \to 0$. The effect of collision loss at depths which are not too small is to cut the spectrum off for energies below the critical energy, and to make the number of electrons finite as $E \to 0$. For low energies therefore the collision loss does not introduce a small correction, but plays an essential part in the process, and hence must enter in some essential way into the solution. In incorporating this feature the solution of this paper differs essentially from all the previous treatments of the problem, and may be regarded as an advance on all the previous work.

. In the present paper the collision loss is treated as a constant loss independent of the energy of the electron. With this assumption a mathematical solution of the cascade equations is given which allows the energy distribution of electrons to be calculated for *all* energies, *including those near and below the critical energy*. The solution is formally in the form of an infinite series, but this series is not strictly a series in powers of β, since β enters essentially into the expression for each term. The first few terms of the solution can be evaluated without undue difficulty, and it can be shown that the main contribution comes from the first term, the second being always smaller than the first, and for all except the very end of a shower less than 30 % of the first, *even for electrons of energy near or less than the critical energy*. Moreover, the solution is valid irrespectively of how small the energy of the electron starting the shower may be, in contrast with the work of the other authors mentioned above. Indeed, for an electron of initial energy less than the critical energy it can be shown that all but the first term of the solution are negligible, so that our method incidentally also gives the complete solution to the problem of the absorption of a low-energy electron by collision loss with a very small accompanying cascade. It also allows one to calculate the energy spectrum of electrons in a shower at small thicknesses, that is, precisely in the region where the usual methods fail. It is found that the energy spectrum at small thicknesses is quite different to what has been usually supposed. We therefore get, for the first time, a quantitative insight into the effect of collision loss on the energy spectrum of cascade electrons at all depths.

For the processes of radiation loss and pair creation, we take the exact cross-sections calculated by Bethe & Heitler valid for extremely high energies where screening is complete. The decrease of these cross-sections at energies so low that screening is incomplete will not be considered. These, or cruder approximations thereto, also underlie all the previous work. For this general problem a rigorous mathematical solution has been given by K. S. K. Iyengar (1942), but his method is not suitable for obtaining numerical results without laborious calculation.*

When the effect of collision loss is negligible, i.e. for energies large compared with the critical energy, it appears that the figures of Bhabha & Heitler are not in general in error by more than 30 %, and give the position of the maximum of the shower correctly. The approximations of Carlson & Oppenheimer make the maximum

---

* An empirical method of allowing for incomplete screening has recently been given by Corben (1941). But his treatment of collision loss follows that of Serber and therefore has the same defects. Corben's approximate method of allowing for incomplete screening could be incorporated into the solution of this paper.

number of particles appear at a depth which is too large by 30 % to 60 %, and give an energy spectrum which is incorrect at small thicknesses. Their figures are in error by much more than 30 % even in the region where collision loss is negligible. On the other hand, the figures given by Landau & Rumer (1938) are often in error by a factor 30, due to the fact that these authors have failed to carry their elegant method through to the end with mathematical rigour. For energies large compared with the critical energy the figures of this paper are correct to within 5 %, and hence are a considerable advance on all the previous work. But even for electrons of energy near or below the critical energy the figures for the energy spectrum of electrons calculated in this paper by taking account of collision loss are accurate for all but the end of a shower to within 30 %, and are hence as accurate as those which have been given hitherto for energies far above the critical energy where collision loss is negligible. Moreover, the present calculations show that even for electrons of the critical energy, the effect of collision loss is so large that the number of electrons may be as low as one-seventh of what was previously supposed.

PHYSICAL ASSUMPTIONS AND THE GENERAL EQUATIONS

The effective differential cross-section for the emission by an electron of energy E of a quantum of energy lying between E' and $E' + dE'$ while passing through the field of an atom of nuclear charge Ze have been calculated by Bethe & Heitler (1934) and can be written in the form

$$4 \frac{Z^2}{137} \left(\frac{e^2}{mc^2} \right)^2 \log 183 Z^{-\frac{1}{3}} \, R(E, E') \frac{E' dE'}{E^2}, \tag{1}$$

where

$$R(E, E') = \left(1 - \frac{4}{3} \frac{E}{E'} + \frac{4}{3} \frac{E^2}{E'^2} \right) (\chi_1 + \chi_2) - \chi_2 \tag{2}$$

and

$$\chi_1 = \frac{\phi_1(\rho) + 4 \log Z^{-\frac{1}{3}}}{4 \log 183 Z^{-\frac{1}{3}}}, \quad \chi_2 = \frac{\phi_1(\rho) - \phi_2(\rho)}{8 \log 183 Z^{-\frac{1}{3}}}, \tag{3}$$

with

$$\rho = \frac{100}{Z^{\frac{1}{3}}} \left| \frac{mc^2 E'}{E(E - E')} \right|, \tag{4}$$

m being the electron mass. Here ϕ_1 and ϕ_2 are functions of ρ only, and are given in the paper of Bethe & Heitler. χ_1 and χ_2 decrease monotonically as ρ increases, and take their largest value when $\rho = 0$, in which case $\chi_1 = 1$ and $\chi_2 = (12 \log 183 Z^{-\frac{1}{3}})^{-1}$. Thus, even for the heaviest atoms χ_2/χ_1 is less than 5 % when $\rho = 0$, and it becomes rapidly smaller as ρ increases, being already less than 1 % even for $\rho = 0.5$. It is completely negligible for $\rho \geqslant 0.8$. Thus, if an accuracy of not more than 5 % is aimed at χ_2 can be neglected altogether. Now $\rho \ll 1$, that is, by (4)

$$\left| \frac{E(E - E')}{E' mc^2} \right| \gg 100 Z^{-\frac{1}{3}} \tag{5}$$

corresponds to the case in which the screening of the nucleus by the surrounding atomic electrons is complete, and then, introducing the values of χ_1 and χ_2 given above for $\rho = 0$, (2) simplifies to

$$R_0(E, E') = 1 - (\tfrac{4}{3} + \alpha)\left(\frac{E}{E'} - \frac{E^2}{E'^2}\right), \tag{6}$$

with
$$\alpha = \frac{1}{9 \log 183 Z^{-\frac{1}{3}}}.$$

This is the expression which will be used in the cascade calculations of this paper. It depends only on the ratio E/E', whereas the general expression (2) depends not only on this ratio but on the absolute magnitudes of E and E'. Moreover, the properties of the atom only affect α which is about 0·06 for lead, and still less for lighter atoms, so that the expression (6) is practically independent of Z. To the accuracy of the calculations of this paper α could equally well be neglected, but we shall not do so since its inclusion presents no difficulty.

The effective differential cross-section for the creation by a quantum of energy E of an electron pair, one particle of which has an energy between E' and $E' + dE'$, while passing through the field of an atom of charge Ze has also been calculated by Bethe & Heitler, and can be written in the form

$$4 \frac{Z^2}{137}\left(\frac{e^2}{mc^2}\right)^2 \log 183 Z^{-\frac{1}{3}} R(E', E) \frac{dE'}{E}, \tag{7}$$

with R given by (2). Here again, for very high energies, more accurately, for the case of complete screening, R takes on the simpler form given by (6), and we shall use this specialized form in the calculations of this paper.

Since in the calculations of this paper we use R_0 instead of R, it is important to know the error which this introduces. R_0 given by (6) is accurate only when the inequality (5) is satisfied. It therefore differs from the more general expression (2) provided $E - E'$ is small enough, however large E may be. But in practice, more detailed considerations show that, except for very small values of $E - E'$ which are not important in our calculations, the difference between R_0 and R is less than 20 % provided E is greater than 30 MeV in lead and 70 MeV in air. Now, as will be seen below, the critical energy at which collision loss becomes important is 7 MeV in lead and 103 MeV in air. Thus the fact that the cross-section (6) is not the exact one for low energies* restricts the accuracy of our calculations to 20 % in lead even for electrons above the critical energy, while in air the error is within about 10 % above the critical energy, and considerably less than 20 % even for particles below the critical energy, since the majority of these are produced by radiation loss and pair creation by particles and quanta above the critical energy for which the formula (6) is accurate to a much greater degree. It should be noticed that whereas the use of (6) instead of (2) even for low energies introduces relatively small errors, the neglect of collision loss completely alters the whole picture near and below the critical energy.

* An approximate allowance for this could be made here by following Corben.

Moreover, as has already been mentioned, (6) is always greater than (2), so that the use of R_0 instead of R, even when the inequality (5) is not satisfied, results in slightly more multiplication than is actually the case.

The probability that in travelling a thickness dx of a substance containing N atoms of nuclear charge Ze per unit volume an electron of energy E emits a quantum of energy E', or a quantum of energy E creates a pair of which one particle has an energy E', is got by multiplying (1) or (7) respectively by $N\,dx$. It is therefore convenient to introduce a quantity l of the dimensions of length, defined by

$$l^{-1} = 4\frac{Z^2N}{137}\left(\frac{e^2}{mc^2}\right)^2 \log 183 Z^{-\frac{1}{3}}. \tag{8}$$

For substances containing several types of atoms one must take on the right-hand side of (8) a sum of similar expressions, one for each type of atom. This quantity l will be known as the characteristic unit of length for that substance. As in all the previous papers, it is convenient to measure all lengths in a given substance in terms of the characteristic length l and denote length so measured by t, so that $t \equiv x/l$ and is a dimensionless quantity. The values of l for different substances are given in the first row of table 1.

TABLE 1

	air	H_2O	Al	Fe	Pb
l in cm.	34,200	43·4	9·80	1·84	0·525
mean collision loss in MeV per cm. $= (-\partial E/\partial x)_{coll.}$	$3·03 \times 10^{-3}$	2·64	5·67	14·1	13·2
β in MeV $= l(-\partial E/\partial x)_{coll.}$	103·0	114·6	55·56	25·88	6·927

The only other physical process which plays an essential part in the cascade theory is the 'ionization' or collision loss of electrons. This loss is a minimum at energies near twice the rest energy of an electron, and increases rapidly for lower energies so that any electron whose energy has fallen as low as twice the rest energy, i.e. about one million electron volts, can be regarded as completely stopped. For higher energies the collision loss increases logarithmically with the energy, and in the whole range from 5 to 150 MeV in which alone the collision loss plays a dominating part, it rises by less than 50 %. It will therefore be treated as a constant loss independent of the energy. Hence, measuring length in the characteristic unit (8), the collision loss is

$$-\left(\frac{\partial E}{\partial t}\right)_{coll.} = -l\left(\frac{\partial E}{\partial x}\right)_{coll} \equiv \beta, \tag{9}$$

where β is a constant independent of the energy. Since the collision loss is very nearly proportional to ZN, while l is proportional to $Z^{-2}N^{-1}$, it follows that β is very nearly proportional to Z^{-1} and is independent of N. It has the dimensions of energy. The second row in table 1 gives the mean values that are taken in this paper for the collision loss $-(\partial E/\partial x)_{coll.}$ in different substances, and the third row then gives the corresponding values of β as defined by (9).

From (1) the total energy loss of an electron due to radiation is

$$-\left(\frac{\partial E}{\partial t}\right)_{\text{rad.}} = E \int_0^E R(E, E') \frac{E'^2 dE'}{E^3} . \tag{10}$$

Using (2) for R, it appears that the integral only increases logarithmically with E, and for very large E, where in fact (2) tends to (6), it takes the constant value $(1 + \frac{1}{2}\alpha) \approx 1$. Thus, for high energies,

$$-\left(\frac{\partial E}{\partial t}\right)_{\text{rad.}} \approx E. \tag{11}$$

The critical energy in a substance is usually defined as the energy at which the radiation loss of an electron is equal to its collision loss. By equating (9) and (11) the critical energy is seen to be very nearly β. For the purposes of this paper, in which the collision loss is treated accurately, it is logical to define the critical energy as exactly equal to β, for this is the physically significant quantity which appears in the equations. The critical energy β has therefore another physically significant interpretation. It is the collision energy loss suffered by an electron in travelling the characteristic unit of length in the substance. Hence an electron whose energy has fallen to the critical energy will at the most travel a distance of one unit before its energy is reduced to zero.

For quanta of energy below 5 MeV in lead and 25 MeV in air, the Compton effect becomes larger than the pair creation. But even for energies down to 5 MeV the angular scattering of the quanta is small, and in consequence in each Compton process a small fraction of the energy of the quantum is given to the electron. The effect of successive Compton processes is therefore to reduce the energy of low-energy quanta in a large number of small steps just as collision loss reduces the energy of electrons, and indeed this energy loss can be taken into account in our theory in the same way. Calculation shows that the resulting energy loss of a quantum per unit length of path traversed is between one-eighth and one-tenth of the collision energy loss of electrons. It can therefore be omitted altogether, since its effect would be to modify the spectrum of quanta of energies below about one-tenth of the critical energy.

Consider a layer of substances, upon the surface of which electrons and quanta with a known energy distribution impinge normally. We wish to know the energy distribution of electrons and quanta at some distance t from the surface of the layer. The problem may be treated as a one-dimensional one as far as calculating the number of particles at a depth t is concerned, since the angular deflexion of a particle or quantum of energy E resulting from a radiation or pair creation process is of the order mc^2/E, and is hence small. Denote by $P(E, t) dE$ the number of particles (electrons and positrons) whose energies lie between E and $E + dE$ at a distance t from the surface of the substance, all distances t being now measured normally from the surface of the layer in the characteristic unit of length l. Let the corresponding expression for the number of quanta be $Q(E, t) dE$. Then, owing to the radiation, pair

creation and collision processes given by (1), (7) and (9), this number varies with the thickness, and the differential equations determining their variation can be written down exactly as in the paper of Landau & Rumer. They are

$$\frac{\partial P(E,t)}{\partial t} = \int_E^\infty P(E',t)\, R(E', E'-E)\frac{(E'-E)\,dE'}{E'^2} - P(E,t)\int_0^E R(E,E')\frac{E'\,dE'}{E^2}$$
$$+ 2\int_E^\infty Q(E',t)\, R(E,E')\frac{dE'}{E'} + \beta\frac{\partial P(E,t)}{\partial E}, \quad (12a)$$

$$\frac{\partial Q(E,t)}{\partial t} = \int_E^\infty P(E',t)\, R(E', E)\frac{E\,dE'}{E'^2} - Q(E,t)\int_0^E R(E',E)\frac{dE'}{E}. \quad (12b)$$

The first two terms on the right-hand side of (12a) give the change in the number of particles of energy E due to radiation loss, the third term that due to pair creation by quanta, and the last term that due to collision loss. Similarly, the first term on the right-hand side of (12b) gives the increase in the number of quanta of energy E due to radiation by electrons and the second term their decrease due to disappearance by pair creation. In these equations R is strictly given by the general expression (2), which depends not only on the ratio E/E' but also on the value of E/mc^2. For this general problem no rigorous solution* can be obtained by the method of this paper and the general method developed by K. S. K. Iyengar must be used. While his method has the advantage of great generality and mathematical rigour, it is not so easy to derive numerical results from it. Moreover, it depends on an ingenious use of the solution neglecting collision loss derived in this paper. But quite apart from this, as the previous discussion shows, it is possible with considerable accuracy to take for R over the whole energy range the form (6) which it assumes for very high energies, and which is only a function of the ratio E/E'. It is then possible to solve the equations (12) by using the Mellin transformation. This method has the advantage that it is suitable for obtaining numerical results.

Introduce two new functions $p(s,t)$ and $q(s,t)$ of a variable s defined by

$$p(s,t) = \int_0^\infty E^{s-1}\, P(E,t)\,dE, \quad (13a)$$

$$q(s,t) = \int_0^\infty E^{s-1}\, Q(E,t)\,dE. \quad (13b)$$

It then follows from the theory of the Mellin transform that

$$P(E,t) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} E^{-s}\, p(s,t)\,ds, \quad (14a)$$

$$Q(E,t) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} E^{-s}\, q(s,t)\,ds, \quad (14b)$$

where σ is such that when $R(s) > \sigma$, $p(s,t)$ and $q(s,t)$ are analytic. The path of the s integration is a line parallel to the imaginary axis and to the right of it.

* An approximate treatment can be given by following Corben.

Following Landau & Rumer, multiply the equations (12) by E^{s-1} and integrate with respect to E from 0 to ∞. The last term on the right-hand side of (12a) then gives

$$\beta \int_0^\infty E^{s-1} \frac{\partial P(E,t)}{\partial E} dE = \beta \left[E^{s-1} P(E,t) \right]_0^\infty - \beta(s-1) \int_0^\infty E^{s-2} P(E,t) dE,$$

on integrating by parts. It is obvious from the physics of the problem that $E^{s-1} P(E,t)$ must vanish at infinity, since initially there are only particles of finite energy. Moreover, when collision loss is taken into account this expression also vanishes at $E = 0$ provided $R(s) > 1$. Thus the term on the left of the above equation just reduces to the second term on the right, which by (13a) can be written as $-\beta(s-1)p(s-1,t)$. The first term on the right of (12a) gives, if we interchange the orders of the E and E' integrations,

$$\int_0^\infty dE' \, P(E',t) \, E'^{s-1} \int_0^{E'} dE \, R(E', E'-E) \frac{(E'-E)E^{s-1}}{E'^{s+1}}.$$

Now if R is a function only of the ratio E/E', as is the case if we take for it the expression (6), then the E integral becomes a function of s only and is independent of the value of E', so that remembering (13) the double integral becomes $p(s,t)$ multiplied by a function of s.* All the other integrals can be treated similarly and the equations reduce to

$$\frac{\partial p(s,t)}{\partial t} = - A_s p(s,t) + B_s q(s,t) - \beta(s-1) p(s-1,t), \tag{15a}$$

$$\frac{\partial q(s,t)}{\partial t} = C_s p(s,t) - D q(s,t), \tag{15b}$$

where A_s, B_s and C_s are functions of s denfied by

$$A_s \equiv \int_0^1 \left\{ \epsilon - (\tfrac{4}{3} + \alpha) \left(1 - \frac{1}{\epsilon} \right) \right\} \{ 1 - (1-\epsilon)^{s-1} \} d\epsilon$$

$$= (\tfrac{4}{3} + \alpha) \left\{ \frac{d}{ds} \log \Gamma(s) + \gamma - 1 + \frac{1}{s} \right\} + \frac{1}{2} - \frac{1}{s(s+1)}, \tag{16a}$$

$$B_s \equiv 2 \int_0^1 \{ 1 - (\tfrac{4}{3} + \alpha)(\epsilon - \epsilon^2) \} \epsilon^{s-1} d\epsilon$$

$$= 2 \left\{ \frac{1}{s} - (\tfrac{4}{3} + \alpha) \frac{1}{(s+1)(s+2)} \right\}, \tag{16b}$$

$$C_s \equiv \int_0^1 \left\{ 1 - (\tfrac{4}{3} + \alpha) \left(\frac{1}{\epsilon} - \frac{1}{\epsilon^2} \right) \right\} \epsilon^s d\epsilon$$

$$= \frac{1}{s+1} + (\tfrac{4}{3} + \alpha) \frac{1}{s(s-1)} \tag{16c}$$

and

$$D \equiv \int_0^1 \{ 1 - (\tfrac{4}{3} + \alpha)(\epsilon - \epsilon^2) \} d\epsilon$$

$$= \tfrac{7}{6} - \tfrac{1}{6}\alpha. \tag{16d}$$

* Owing to the form of R_0 the E integral diverges when $E = E'$, and the first and second integrals in (12a) have to be taken together in the Cauchy sense to give a convergent result. These together then give $-A_s p(s,t)$.

γ is the Euler-Mascheroni constant. The equations (15) were given by Landau & Rumer, but with the important omission of the last term in (15a) representing the effect of collision loss.*

Differentiating (15a) with respect to t and then eliminating $q(s,t)$ by the use of (15b) and then (15a) we obtain the second-order equation

$$\frac{\partial^2}{\partial t^2} p(s,t) + (A_s + D) \frac{\partial}{\partial t} p(s,t) + (A_s D - B_s C_s) p(s,t)$$
$$= -\beta(s-1)\left\{\frac{\partial}{\partial t} + D\right\} p(s-1,t). \qquad (17)$$

From the definitions (13) it is obvious that $p(2,t)$ gives the total energy carried by particles and $q(2,t)$ that carried by quanta at the depth t, while $p(1,t)$ gives the total number of particles at this depth. Adding (15a) and (15b) and putting $s = 2$, we get

$$\frac{\partial}{\partial t}\{p(2,t) + q(2,t)\} = (C_2 - A_2) p(2,t) + (B_2 - D) q(2,t) - \beta p(1,t).$$

Since the left-hand side of the above equation gives the rate of increase of the total energy of the shower with depth, and the radiation and pair creation processes do not change the total energy of the shower, the terms independent of β on the right-hand side of the above equation must vanish irrespective of the values of $p(2,t)$ and $q(2,t)$. It follows that the coefficients of these terms must vanish, and hence that $A_2 = C_2$ and $B_2 = D$, relations which can be verified directly from (16). The above equation therefore reduces to

$$\frac{\partial}{\partial t}\{p(2,t) + q(2,t)\} = -\beta p(1,t),$$

which expresses the physically evident result that the rate of diminution of the total energy of a shower is equal to the rate of energy loss by collision. This equation also shows that, since the left-hand side is always finite, *the total number of particles $p(1,t)$ at a thickness t must always be finite when collision loss is taken into account*. This is not so when collision loss is neglected ($\beta = 0$).

CASCADES WITHOUT COLLISION LOSS

It is first necessary to give a complete treatment of the cascade problem neglecting collision loss, that is, the solution of the equations (12), with $\beta = 0$, since the work of the next section is based upon it.

As has been shown by Landau & Rumer, the general solution of the equations

---

* With the approximations made by Carlson & Oppenheimer, we should have $A_s = 2(1 - 1/s)$, $B_s = 4/3s$, $C_s = 1/(s-1)$, $D = 2/3$. Hence A_s goes seriously wrong for $s \sim 1$. Although their values of λ_s are not in general in error by more than 50 %, since λ appears in (32) in the form $\exp(-\lambda_s t)$, the resulting error in the number of particles may be very large. See footnote on p. 283.

(15) when $\beta = 0$, and consequently of equation (17) neglecting the right-hand side, is of the form

$$p(s, t) = a_s e^{-\lambda_s t} + b_s e^{-\mu_s t}, \qquad (18)$$

with a similar expression for $q(s, t)$, where a_s, b_s are arbitrary functions of s independent of t which have to be chosen to satisfy the boundary conditions, and λ_s and μ_s are the two roots of the quadratic

$$X^2 - (A_s + D) X + (A_s D - B_s C_s) = 0,$$

whence $\lambda_s + \mu_s = A_s + D$, a relation which will often be used later, and

$$\lambda_s = \tfrac{1}{2}(A_s + D) - \tfrac{1}{2}\sqrt{\{(A_s - D)^2 + 4 B_s C_s\}}, \qquad (19a)$$

$$\mu_s = \tfrac{1}{2}(A_s + D) + \tfrac{1}{2}\sqrt{\{(A_s - D)^2 + 4 B_s C_s\}}. \qquad (19b)$$

It follows from the definitions (19) that $\mu_s > \lambda_s$ when s is real and greater than 1.

Certain general results can be deduced at once from the form of the solutions (18). Since, as was shown at the end of the last section, $A_2 = C_2$ and $B_2 = D$, hence $A_2 D - B_2 C_2 = 0$, and the smaller of the roots λ and μ must vanish at $s = 2$. Hence $\lambda_s = 0$ when $s = 2$ and this is a consequence of the conservation of energy. It further follows from (19b) and (16) that

$$\mu_2 = C_2 + D = 1 + \tfrac{1}{2}\alpha + \tfrac{7}{9} - \tfrac{1}{6}\alpha = \tfrac{16}{9} + \tfrac{1}{3}\alpha.$$

Putting $s = 2$ in (18) we see that the total energy of the particles or quanta in a shower tends to a constant value since $\lambda_2 = 0$ and the terms proportional to $\exp(-\mu_2 t)$ become negligible for $t \gg 1/\mu_2 = 9/(16 + 3\alpha)$. Thus, at large thicknesses, $\partial q(2, t)/\partial t$ vanishes, and it follows from (15b) that

$$\frac{p(2, t)}{q(2, t)} = \frac{D}{C_2} = \frac{14 - 3\alpha}{18 + 9\alpha} \approx \frac{7}{9}.$$

Hence, after a thickness of about one unit, the ratio of the total energy carried by the particles to that carried by the quanta reaches a constant value, the quanta carrying somewhat more than half the total energy of the shower. The energy is of course carried by progressively more and more particles and quanta of lower energy. This conclusion is altered if collision loss is taken into account, the energy carried by the particles then being relatively less, especially at large thicknesses. The above discussion also shows that the effect of conditions at the boundary does not make itself felt for more than a distance of about one unit. This is the average range in which 'transition effects' must take place when the shower passes from one substance to another.

Now consider a shower started by one electron or positron of energy E_0 entering the surface of the substance unaccompanied by any quanta. The boundary conditions at $t = 0$ are then

$$P(E, 0) = \delta(E - E_0), \quad Q(E, 0) = 0. \qquad (20)$$

From this, using (13), it follows that at $t = 0$

$$p(s, 0) = E_0^{s-1}, \tag{21a}$$

$$q(s, 0) = 0, \tag{21b}$$

and from (15a)
$$\left\{\frac{\partial}{\partial t} p(s, t)\right\}_{t=0} = -A_s E_0^{s-1}. \tag{22}$$

We have therefore to find a solution of equations (15) satisfying the boundary conditions (21) or of equation (17) with the boundary conditions (21a) and (22).

The conditions (21a) and (22) determine a_s and b_s, and give finally

$$p(s, t) = E_0^{s-1}\left\{\frac{D-\lambda}{\mu-\lambda} e^{-\lambda t} + \frac{\mu-D}{\mu-\lambda} e^{-\mu t}\right\}, \tag{23}$$

where for brevity we have omitted the suffix s. Owing to singularities in A_s, B_s or C_s, λ_s and μ_s are singular when $s = 1$, 0 or a negative integer. It should be noticed at once that, although λ_s and μ_s are not one-valued functions of s due to the appearance of a square root in their definition, the functions $p(s, t)$ and $q(s, t)$ are symmetrical in the two roots λ and μ and are therefore one-valued everywhere in the s plane, although a part of them, for example the part containing $e^{-\lambda t}$, is by itself not one-valued. This point is of importance in the general theory of the next section.

The number of particles of energy E at the depth t namely $P(E, t)$ can be calculated at once by using (14a) and is

$$P(E, t) = \frac{1}{2\pi i E_0} \int_{\sigma-i\infty}^{\sigma+i\infty} \left(\frac{E_0}{E}\right)^s \left\{\frac{D-\lambda}{\mu-\lambda} e^{-\lambda t} + \frac{\mu-D}{\mu-\lambda} e^{-\mu t}\right\} ds. \tag{24}$$

(24) is the *exact* solution of the cascade equations for a shower started by an incident particle of energy E_0 if collision loss is neglected.

To proceed further it is necessary to know the dependence of A, λ and μ on s. This can be calculated at once by using (16) and (19). The values of λ and μ are given in table 2 for real values of s from 1 to 6 at intervals of 0.1. The asymptotic form of these functions for large s will also be required. From (16) and (19) it follows that asymptotically for large s

$$A_s \approx \alpha'\left\{\log s + \gamma - 1 + \frac{1}{2s}\right\} + \tfrac{1}{2}, \tag{25a}$$

$$\lambda_s \approx D - \frac{B_s C_s}{A_s - D} \approx D - \frac{2}{\alpha' s(s+1) \log s}, \tag{25b}$$

$$\mu_s \approx A_s + \frac{B_s C_s}{A_s - D} \approx A_s + \frac{2}{\alpha' s(s+1) \log s}, \tag{25c}$$

where we have written $\alpha' = \tfrac{1}{2} + \alpha$ for brevity.

TABLE 2

s	λ_s	μ_s	$\log\left(\dfrac{D-\lambda_s}{\mu_s-\lambda_s}\right)$	λ'_s	$\dfrac{d}{ds}\log\left(\dfrac{D-\lambda}{\mu-\lambda}\right)$	λ''_s	$\dfrac{d^2}{ds^2}\log\left(\dfrac{D-\lambda}{\mu-\lambda}\right)$
1·0	$-\infty$	$+\infty$					
1·1	$-3\cdot787$	$+4\cdot712$					
1·2	$-2\cdot279$	$3\cdot339$	$-0\cdot6102$	$9\cdot450$	$+0\cdot043$	$-65\cdot42$	$-1\cdot34$
1·3	$-1\cdot569$	$2\cdot749$	$-0\cdot6120$	$5\cdot410$	$-0\cdot070$	$-24\cdot67$	$-0\cdot941$
1·4	$-1\cdot125$	$2\cdot413$	$-0\cdot6225$	$3\cdot653$	$-0\cdot152$	$-12\cdot47$	$-0\cdot778$
1·5	$-0\cdot8121$	$2\cdot200$	$-0\cdot6414$	$2\cdot691$	$-0\cdot225$	$-7\cdot400$	$-0\cdot699$
1·6	$-0\cdot5751$	$2\cdot054$	$-0\cdot6672$	$2\cdot091$	$-0\cdot294$	$-4\cdot855$	$-0\cdot665$
1·7	$-0\cdot3876$	$1\cdot951$	$-0\cdot7005$	$1\cdot683$	$-0\cdot359$	$-3\cdot422$	$-0\cdot646$
1·8	$-0\cdot2348$	$1\cdot877$	$-0\cdot7400$	$1\cdot388$	$-0\cdot423$	$-2\cdot539$	$-0\cdot639$
1·9	$-0\cdot1079$	$1\cdot824$	$-0\cdot7844$	$1\cdot165$	$-0\cdot487$	$-1\cdot960$	$-0\cdot631$
2·0	$-0\cdot0000$	$1\cdot786$	$-0\cdot8365$	$0\cdot9904$	$-0\cdot550$	$-1\cdot561$	$-0\cdot621$
2·1	$+0\cdot0915$	$1\cdot760$	$-0\cdot8944$	$0\cdot8493$	$-0\cdot610$	$-1\cdot275$	$-0\cdot600$
2·2	$+0\cdot1706$	$1\cdot743$	$-0\cdot9577$	$0\cdot7330$	$-0\cdot669$	$-1\cdot060$	$-0\cdot575$
2·3	$+0\cdot2389$	$1\cdot734$	$-1\cdot029$	$0\cdot6357$	$-0\cdot725$	$-0\cdot8940$	$-0\cdot537$
2·4	$+0\cdot2982$	$1\cdot731$	$-1\cdot103$	$0\cdot5531$	$-0\cdot775$	$-0\cdot7615$	$-0\cdot490$
2·5	$+0\cdot3499$	$1\cdot733$	$-1\cdot182$	$0\cdot4825$	$-0\cdot821$	$-0\cdot6540$	$-0\cdot433$
2·6	$0\cdot3950$	$1\cdot740$	$-1\cdot268$	$0\cdot4218$	$-0\cdot861$	$-0\cdot5639$	$-0\cdot372$
2·7	$0\cdot4345$	$1\cdot750$	$-1\cdot355$	$0\cdot3693$	$-0\cdot896$	$-0\cdot4882$	$-0\cdot306$
2·8	$0\cdot4689$	$1\cdot763$	$-1\cdot446$	$0\cdot3237$	$-0\cdot921$	$-0\cdot4231$	$-0\cdot226$
2·9	$0\cdot4995$	$1\cdot778$	$-1\cdot539$	$0\cdot2841$	$-0\cdot944$	$-0\cdot3672$	$-0\cdot174$
3·0	$0\cdot5261$	$1\cdot795$	$-1\cdot635$	$0\cdot2500$	$-0\cdot959$	$-0\cdot3220$	$-0\cdot115$
3·1	$0\cdot5496$	$1\cdot815$	$-1\cdot683$	$0\cdot2204$	$-0\cdot967$	$-0\cdot2769$	$-0\cdot0585$
3·2	$0\cdot5703$	$1\cdot835$	$-1\cdot734$	$0\cdot1945$	$-0\cdot970$	$-0\cdot2405$	$-0\cdot0095$
3·3	$0\cdot5886$	$1\cdot857$	$-1\cdot786$	$0\cdot1721$	$-0\cdot968$	$-0\cdot2094$	$+0\cdot0354$
3·4	$0\cdot6048$	$1\cdot879$	$-1\cdot840$	$0\cdot1527$	$-0\cdot964$	$-0\cdot1819$	$+0\cdot0676$
3·5	$0\cdot6192$	$1\cdot903$	$-1\cdot895$	$0\cdot1355$	$-0\cdot955$	$-0\cdot1585$	$+0\cdot0987$
3·6	$0\cdot6321$	$1\cdot926$					
3·7	$0\cdot6436$	$1\cdot950$					
3·8	$0\cdot6536$	$1\cdot975$					
3·9	$0\cdot6628$	$1\cdot999$					
4·0	$0\cdot6710$	$2\cdot023$					
4·1	$0\cdot6785$	$2\cdot048$					
4·2	$0\cdot6851$	$2\cdot072$					
4·3	$0\cdot6910$	$2\cdot097$					
4·4	$0\cdot6969$	$2\cdot121$					
4·5	$0\cdot7019$	$2\cdot145$					
4·6	$0\cdot7061$	$2\cdot169$					
4·7	$0\cdot7102$	$2\cdot193$					
4·8	$0\cdot7143$	$2\cdot216$					
4·9	$0\cdot7176$	$2\cdot239$					
5·0	$0\cdot7207$	$2\cdot262$					
5·1	$0\cdot7239$	$2\cdot284$					
5·2	$0\cdot7266$	$2\cdot307$					
5·3	$0\cdot7289$	$2\cdot329$					
5·4	$0\cdot7318$	$2\cdot351$					
5·5	$0\cdot7335$	$2\cdot372$					
5·6	$0\cdot7356$	$2\cdot394$					
5·7	$0\cdot7372$	$2\cdot415$					
5·8	$0\cdot7389$	$2\cdot436$					
5·9	$0\cdot7409$	$2\cdot456$					
6·0	$0\cdot7424$	$2\cdot480$					

It is convenient, as in the paper by Bhabha & Heitler, to introduce a variable y defined by

$$y = \log \frac{E_0}{E}. \tag{26}$$

It should be noted at once that when $E > E_0$ (24) vanishes exactly. For when $E > E_0$, y is negative, and the path of the s integration in (24) can be deformed into an infinite semicircle to the right with the point σ as centre. The expression in curly brackets tends to zero on this semicircle, as also the term outside it, namely, $\exp(-|y|s)$, so that (24) vanishes. This is, of course, necessary for the physical interpretation.

Now consider (24) when $E < E_0$. Since $\lambda_s < \mu_s$ for real $s > 1$, as is also shown clearly by table 2, it follows that, except for very small t, the main contributions to (24) come from the first term in curly brackets having λ_s in the exponent. Write $P(E, t)$ given by (24) as the sum of two terms P_λ and P_μ defined by

$$P_\lambda \equiv \frac{1}{2\pi i E_0} \int_{\sigma-i\infty}^{\sigma+i\infty} \left(\frac{E_0}{E}\right)^s \frac{D-\lambda}{\mu-\lambda} e^{-\lambda t} ds, \tag{27}$$

$$P_\mu \equiv \frac{1}{2\pi i E_0} \int_{\sigma-i\infty}^{\sigma+i\infty} \left(\frac{E_0}{E}\right)^s \frac{\mu-D}{\mu-\lambda} e^{-\mu t} ds. \tag{28}$$

It will now be shown that (27) can be evaluated by the saddle-point method with an error well within 5 %. Introduce a function ψ such that e^ψ is equal to the integrand of (27). Then, remembering (26),

$$\psi(s) \equiv ys - \lambda_s t + \log \frac{D-\lambda_s}{\mu_s - \lambda_s}. \tag{29}$$

Since $C_s \to \infty$ as $s \to 1$ it follows that $\lambda_s \to -\infty$ and hence $\psi \to \infty$ as $s \to 1$. Moreover, it follows from (25) that $\psi \to \infty$ as $s \to \infty$. Hence ψ must have a minimum as s increases along the real axis from 1 to ∞. Denote the particular value of s, where

$$\frac{\partial \psi}{\partial s} = 0$$

by s_0. We shall call this the saddle-point. Hence, differentiating (29), s_0 is determined by

$$y - \lambda'_{s_0} t + \left(\frac{d}{ds} \log \frac{D-\lambda_s}{\mu_s - \lambda_s}\right)_{s=s_0} = 0. \tag{30}$$

A dash affixed to a symbol will be used to denote differentiation with respect to s. The value of s_0 depends on the values of y and t. Now shift the contour of the integration in (27) to a parallel contour to the right or left so as to make it pass through the point s_0, in other words, take $\sigma = s_0$. This can always be done since the only restriction on σ is $\sigma > 1$. Writing $i\tau \equiv s - s_0$, by Taylor's theorem

$$\psi(s) = \psi(s_0) - \frac{\tau^2}{2} \psi''(s_0) + \dots. \tag{31}$$

We now make the usual approximation of the saddle-point method and replace ψ by the first three terms of its Taylor expansion, namely, the two terms on the right of (31). Then

$$P_\lambda(E,t) = \frac{1}{2\pi E_0} \int_{-\infty}^{\infty} e^{\psi(s_0) - \frac{1}{2}\tau^2 \psi''(s_0)} d\tau$$

$$= \frac{e^{\psi(s_0)}}{E_0 \sqrt{\{2\pi \psi''(s_0)\}}} = \frac{1}{E_0} \frac{e^{ys_0 - \lambda_{s_0}t}}{\sqrt{\{2\pi \psi''(s_0)\}}} \frac{D - \lambda_{s_0}}{\mu_{s_0} - \lambda_{s_0}}, \qquad (32a)$$

where, from (30) $\qquad \psi''(s_0) = -\lambda''_{s_0}t + \left(\frac{d^2}{ds^2}\log\frac{D-\lambda_s}{\mu_s - \lambda_s}\right)_{s=s_0}.$ $\qquad (32b)$

It should be noticed that $\psi''(s_0)$ contains t explicitly but not y.

The approximation in deriving (32a) consists in replacing $\exp\{\psi(s_0 + i\tau)\}$ by $\exp\{\psi(s_0) - \tau^2\psi''(s_0)/2\}$. Now, owing to the symmetry of the contour of integration in $\pm i\tau$, it is only the real part of $\exp\psi(s_0 + i\tau)$ which makes a non-vanishing contribution to the integral (27). Hence to check the accuracy of the approximation, it is necessary to compare the values of

$$R\{\exp\psi(s_0 + i\tau)\}\exp\{-\psi(s_0)\} \quad \text{with} \quad \exp\{-\tau^2\psi''(s_0)/2\}.$$

This has been done for two examples, namely, $y = 4, t = 10$ and $y = 10, t = 10$, the corresponding values of s_0 being 2·84 and 2·03 respectively. The figures are given in table 3 for different values of τ, which shows that the error introduced by evaluating (27) by the saddle-point method is about 2 %. It appears, therefore, that the accuracy of (32) depends on an accurate determination of the saddle-point s_0 from equation (30). λ'_s and $\frac{d}{ds}\left(\log\frac{D-\lambda}{\mu-\lambda}\right)$ are given in the fifth and sixth columns of table 2. Since s is given at intervals of 0·1 in the table, s_0 has been determined by interpolation by using Newton's formula. The method is described in another paper by Chakrabarty (1942). In the seventh and eighth columns of table 2 are also given the values of λ''_s and $\frac{d^2}{ds^2}\left(\log\frac{D-\lambda}{\mu-\lambda}\right)$, which are required for calculating (32b).

TABLE 3. ACCURACY OF THE SADDLE-POINT METHOD

| | $y = 4, t = 10$ | | $y = 10, t = 10$ | |
| | $s_0 = 2\cdot84, \psi(s_0) = 1\cdot06, \psi''(s_0) = 3\cdot789$ | | $s_0 = 2\cdot03, \psi(s_0) = 9\cdot16, \psi''(s_0) = 14\cdot06$ | |
τ	$R\{\exp\psi(s_0 + i\tau)$ $-\exp\psi(s_0)\}$	$\exp\{-\frac{1}{2}\tau^2\psi''(s_0)\}$	$R\{\exp\psi(s_0 + i\tau)$ $-\exp\psi(s_0)\}$	$\exp\{-\frac{1}{2}\tau^2\psi''(s_0)\}$
0·0	1·00	1·00	1·00	1·00
0·1	0·961	0·981	0·934	0·932
0·2	0·925	0·927	0·753	0·755
0·4	0·744	0·739	0·337	0·325
0·6	0·529	0·506	0·078	0·078
0·8	0·305	0·298	0·009	0·011

Since inaccuracies in the determination of the saddle-point s_0 cause large altera-
tions in the result (32a), the logarithmic term in (29) can in no circumstances be
neglected, as has been done by Landau & Rumer. Nor have these authors used the
reverse of the Mellin transform equation (14a) for calculating $P(E,t)$, so that their
numerical figures are in error by large factors, being sometimes thirty times larger
than those of this paper.

The second part P_μ given by (28) cannot be evaluated in the same manner. For-
tunately its order of magnitude can be accurately estimated, from which it appears
that it is small compared with P_λ except for very small t. To estimate the order of
magnitude of the integral (28) consider the path of integration moved to the right
so that $\sigma \gg 1$. Then, everywhere on the contour we can approximately use the asymp-
totic forms of the functions A, λ and μ for large s given by (25). For large s,
$(\mu - D)/(\mu - \lambda) \to 1$, and neglecting terms of order $1/s^2$ in the exponent, we get

$$P_\mu(E,t) \approx \frac{1}{2\pi i E_0} \int_{\sigma-i\infty}^{\sigma+i\infty} e^{ys-\mu_s t}\, ds$$

$$\approx \frac{e^{-(\alpha'(\gamma-1)+\frac{1}{2})t}}{2\pi i E_0} \int_{\sigma-i\infty}^{\sigma+i\infty} \exp\left\{ys - \frac{\alpha' t}{2s}\right\} s^{-\alpha' t}\, ds.$$

Writing s for ys the integral can at once be transformed into the well-known expres-
sion for a Bessel function (Whittaker & Watson 1927, p. 359) and we get

$$P_\mu(E,t) \approx \frac{1}{E_0} e^{(\alpha'-\gamma\alpha'-\frac{1}{2})\varkappa} \left(\frac{2y}{\alpha' t}\right)^{\frac{1}{2}(\alpha' t-1)} J_{\alpha' t-1}\{\sqrt{(2\alpha' t y)}\}. \qquad (33)$$

Calculation shows that the contribution of P_μ is comparable with P_λ only for small
t, as we should expect. For $t > 4$ it is less than 2 % of P_λ and may therefore be neglected
altogether. A method of calculating (24) when t is small is given in the last section
of this paper.

To get the total number of particles $N(E,t)$ whose energy is greater than E, one
must integrate (24) with respect to E from E to infinity. By interchanging the order
of the E and s integrations the E integration can be carried out first, giving

$$N(E,t) = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \left(\frac{E_0}{E}\right)^{s-1} \frac{1}{s-1} \left\{\frac{D-\lambda}{\mu-\lambda} e^{-\lambda t} + \frac{\mu-D}{\mu-\lambda} e^{-\mu t}\right\} ds. \qquad (34)$$

The integral (34) can be evaluated in the same way as (24). The only difference is
that on the right of (29) ys is replaced by $y(s-1)$ and $-\log(s-1)$ is added to it.
Consequently, for determining the saddle-point, (30) must be replaced by

$$y - \lambda'_{s_0} t + \left(\frac{d}{ds} \log \frac{D-\lambda_s}{\mu_s - \lambda_s}\right)_{s=s_0} - \frac{1}{s_0-1} = 0. \qquad (35)$$

To calculate the maximum value of $N(E,t)$ for a given E, and the value of t at
which this maximum occurs, it is only necessary to consider the part in the curly

brackets in (34) which is proportional to $e^{-\lambda t}$, since the other part is negligible at these thicknesses. The maximum of N therefore occurs when

$$\frac{\partial}{\partial t} N(E,t) = -\frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} e^{y(s-1)} \frac{1}{s-1} \frac{D-\lambda_s}{\mu_s-\lambda_s} \lambda_s e^{-\lambda_s t} \, ds = 0. \tag{36}$$

This integral can also be calculated by the saddle-point method and it will vanish when $\lambda_s = 0$. Now $\lambda_s = 0$ when $s = 2$. Hence, for a given y, the maximum will occur at such a value of t that the corresponding value of s_0 is 2. Calling this value t_m, it follows from (35) that[*]

$$t_m = \frac{1}{\lambda'_2} \left\{ y + \left(\frac{d}{ds} \log \frac{D-\lambda}{\mu-\lambda} \right)_{s=2} - 1 \right\} = 1 \cdot 01 y - 1 \cdot 57. \tag{37}$$

Evaluating (34) by the saddle-point method just like (32a), and then putting $s_0 = 2$ and substituting for t_m by (37), the value of N at this point is

$$N_{\max.}(E) = \frac{e^y}{\sqrt{\{2\pi(\psi''(2)+1)\}}} \frac{D}{\mu_2}$$

$$= 0 \cdot 137 \frac{E_0}{E} \left\{ \log \frac{E_0}{E} - 1 \cdot 31 \right\}^{-\frac{1}{2}}. \tag{38}$$

(38) justifies the formula obtained empirically by Bhabha & Heitler, namely, $N_{\max.} = 0 \cdot 12 (E_0/E)^{0 \cdot 93}$. The lower power of the exponent in their formula has the effect of roughly accounting for the variation of the logarithmic term in the denominator of (38). It appears that (38) is less than Bhabha & Heitler's approximate formula by about 30 %, confirming the estimate of the error in their calculations made by these authors. The numerical factors in (38) are of importance, for they show that even at the maximum, the number of particles is about a tenth of E_0/E, that is, the number that would be obtained if the whole primary energy E_0 were divided equally among particles of energy E. This is easy to understand. It is because there are a number of particles and quanta with energy greater than E, while at the same time a lot of the energy has already been degraded into particles and quanta of energy lower than E.

It should be noticed that according to (30) s_0 increases for a fixed y, as t increases, and the preceding discussion shows that $s_0 = 2$ at the thickness at which the maximum number of particles occur for this y. Hence the saddle-point lies at $s_0 < 2$ before the shower has reached its maximum, and at $s_0 > 2$ after the maximum.

It is interesting to note that $p(s,t)$ given by (23) and $q(s,t)$ are finite for all real values of $s > 1$ but tend to ∞ as $s \to 1$. Hence, writing $s = 1 + \delta$ it follows from (13) that

$$p(1+\delta, t) = \int_0^\infty E^\delta P(E,t) \, dE$$

---

[*] With the approximations of Carlson & Oppenheimer (see footnote on p. 276) equation (37) is replaced by
$$t_m = 1 \cdot 25 y - 1 \cdot 92.$$
This equation makes the maximum lie at a depth which is too great by 25–40 %.

is convergent for $\delta > 0$, but becomes infinite for $\delta = 0$. Hence $P(E, t)$ given by (14) must tend to infinity as $1/E$ for small E. Thus the spectrum of particles and quanta is of the form dE/E for very low energies. This is the result mentioned in the introduction, that if collision loss be neglected then the spectrum of electrons is of the form dE/E for small E. Collision loss completely alters the low-energy spectrum of electrons, as will be seen in the next section. The singularities in $p(s, t)$ and $q(s, t)$ at $s = 1$ are due entirely to the singularity in C at this point, as shown by (16). It can be seen quite easily that this singularity in C comes from the fact that the spectrum of quanta emitted by an electron as given by (1) is of the form dE/E for small E.

The creation of showers started by a quantum instead of by an electron can be treated by the same method. In the boundary condition (20) $P(E, 0)$ and $Q(E, 0)$ are now interchanged, from which it follows that $p(s, 0)$ and $q(s, 0)$ are interchanged in (21), so that the coefficients of the terms proportional to $e^{-\lambda t}$ and $e^{-\mu t}$ are different functions of s. It can be easily shown that a result of this is to replace $(D - \lambda)/(\mu - \lambda)$ by $B/(\mu - \lambda)$ in (27). Since the maximum number of particles occurs when $s_0 = 2$, and at this point $\lambda = 0$ and $B_2 = D$, it follows from (32a) that the number of particles at the maximum is very nearly the same, the difference being due only to the difference in $\psi''(2)$ in the denominator of (32a). It can also be shown that the maximum of $N(E, t)$ always occurs at a thickness which is 0·82 unit greater than the corresponding thickness for electron excited showers as given by (37). Detailed calculations have been carried out in a paper by Chakrabarty, and numerical results are given there. They show, as we should expect, that the curves giving the number of particles as a function of the thickness t are practically the same for a shower produced by an electron or a quantum of the same initial energy, with the difference that the latter curve is shifted to greater thickness by approximately one unit of length over the whole of its range.

<center>CASCADES WITH COLLISION LOSS</center>

We now proceed to solve the equations (15) when $\beta \neq 0$. Consider a shower started by an electron of energy E_0. The boundary conditions at $t = 0$ are then (20), from which (21) can be deduced again. Substituting (21) into (15a) the boundary condition (22) is now replaced by

$$\left\{\frac{\partial}{\partial t} p(s, t)\right\}_{t=0} = - E_0^{s-1}\left\{A_s + \frac{\beta}{E_0}(s - 1)\right\}. \tag{39}$$

The problem can therefore be reduced to the solution of the second order equation (17) with the boundary conditions (21a) and (39) at $t = 0$.

One point should be noted immediately. Since $p(s, t)$ is proportional to E_0^{s-1} and on the right of (17) we have $p(s - 1, t)$ it follows that in effect β only appears in equation (17) and the boundary condition (39) in the ratio β/E_0. Thus, if β be taken as the unit of energy in every substance just as l has been taken as the unit of length, then the equations and the solutions are the same in all substances for initial energies

which are the same multiples of β in each case. Just as there is a transformation of the scale of length in going from one substance to another, so there is a transformation of the scale of energy. In mathematical terms *the number of particles with an energy between $\beta\epsilon$ and $\beta(\epsilon+d\epsilon)$ produced by a primary of energy $\beta\epsilon_0$ at any given thickness t is the same in all substances.*

Since the right-hand side of (17) is smaller than the left by a factor of the order β/E_0 it would appear possible to solve (17) by a perturbation method when $E_0 \gg \beta$. This would give for $p(s,t)$ a solution of the form

$$p(s,t) = E_0^{s-1}\left[\sum_{n=0}^{\infty}\left(\frac{\beta}{E_0}\right)^n \{v_n(s)\,e^{-\lambda_s-nt} + w_n(s)\,e^{-\mu_s-nt}\}\right], \tag{40}$$

where $v_n(s)$ and $w_n(s)$ are functions of s which must also depend on β in order that the boundary conditions (21a) and (39) may be satisfied. A solution of the type (40) is nevertheless useless for getting numerical results, for in calculating $P(E,t)$ from it by equation (14a), the nth term of (40) leads to an integral of the type

$$\frac{1}{2\pi i E_0}\,\beta^n\int_{\sigma-i\infty}^{\sigma+i\infty}\frac{E_0^{s-n}}{E^s}\,\{v_n(s)\,e^{-\lambda_s-nt} + w_n(s)\,e^{-\mu_s\cdot nt}\}\,ds.$$

Writing s for $s-n$ this can be written as

$$\frac{1}{2\pi i E_0}\left(\frac{\beta}{E}\right)^n\int_{\sigma-n-i\infty}^{\sigma-n+i\infty}\left(\frac{E_0}{E}\right)^s \{v_n(s+n)\,e^{-\lambda_s t} + w_n(s+n)\,e^{-\mu_s t}\}\,ds, \tag{41}$$

and since the path of integration can be moved to the right it follows that we can replace $\sigma-n$ by σ in the limits of the above integral. In this integral n only appears in the functions v_n and w_n which do not alter the order of magnitude of the integral. Thus in reality (41) leads to a solution for $P(E,t)$ which is an infinite series in powers of $(\beta/E)^n$ in which the coefficients do not vary rapidly with n. Indeed, it can easily be shown that

$$P(E,t) = \sum_{n=0}^{\infty}\left(-\frac{\beta}{E}\right)^n \frac{1}{2\pi i E_0}\int_{\sigma-i\infty}^{\sigma+i\infty} E^{-s}\frac{\Gamma(s+n)}{\Gamma(s)}$$
$$\times\left\{\frac{(D-\lambda_s)^n\,v(s)\,e^{-\lambda_s t}}{(\lambda_s-\lambda_{s+1})\dots(\lambda_s-\lambda_{s+n})(\lambda_s-\mu_{s+1})\dots(\lambda_s-\mu_{s+n})}\right.$$
$$\left. +\frac{(D-\mu_s)^n\,w(s)\,e^{-\mu_s t}}{(\mu_s-\lambda_{s+1})\dots(\mu_s-\lambda_{s+n})(\mu_s-\mu_{s+1})\dots(\mu_s-\mu_{s+n})}\right\}ds \tag{42}$$

is an exact formal solution of equation (17) where $v(s)$ and $w(s)$ are arbitrary functions of s which have to be chosen to satisfy the boundary conditions (21a) and (39). They can be given as series in powers of β/E_0, the first terms of the series being respectively $E_0^{s-1}(D-\lambda_s)/(\mu_s-\lambda_s)$ and $E_0^{s-1}(\mu_s-D)/(\mu_s-\lambda_s)$, agreeing with the form they have in (24). If the series for $v(s)$ and $w(s)$ be inserted in (42) and the terms properly bracketed to give a strict expansion in powers of β, then, as has been shown by Iyengar, the ensuing series is absolutely convergent. The series (42) as written

is however divergent *for all values of E* due to the appearance of $\Gamma(s+n)$, and as a result the series for $v(s)$ and $w(s)$ are also divergent. Thus the series (42) is strictly not a proper solution at all. Nor is it possible to look upon it as an asymptotic solution of equation (17) valid for small β, for the successive integrals actually increase as one goes from one term to the next when t is large. The solution (42) is therefore quite useless for obtaining any numerical results especially for $E \lesssim \beta$. The methods of solution followed by Snyder and Serber also lead for purposes of calculating the spectrum to series of the type (42) and indeed, Serber's solution can be transformed exactly into (42).

We shall therefore try and find a solution of the equation (17) by generalizing (24), which is an exact solution of this equation when $\beta = 0$. Assume that an exact solution of (17) when $\beta \neq 0$ can be written in the form

$$P(E,t) = \frac{1}{2\pi i E_0} \int_{\sigma-i\infty}^{\sigma+i\infty} \left\{ \frac{E_0}{E+\beta g(s,t)} \right\}^s f(s,t,\beta)\, ds, \qquad (43)$$

where $g(s,t)$ and $f(s,t,\beta)$ are functions of s and t, but not of E. Assume g to be independent of β. Then f must be a function of β, and as will appear presently, is expressible as a power series in β. We assume that σ is such that for $R(s) > \sigma$, $g(s,t)$ and $f(s,t,\beta)$ have no singularities.

Assume further that $f(s,t,\beta)$ satisfies the same boundary conditions at $t = 0$ as the corresponding function in curly brackets of (24), namely

$$f(s,0,\beta) = 1, \quad \left\{ \frac{\partial}{\partial t} f(s,t,\beta) \right\}_{t=0} = -A_s. \qquad (44)$$

It can then be shown that in order that (43) should satisfy the boundary conditions (20), $g(s,t)$ must satisfy at $t = 0$ the boundary conditions

$$g(s,0) = 0, \quad \left\{ \frac{\partial}{\partial t} g(s,t) \right\}_{t=0} = 1. \qquad (45)$$

To see that (43) satisfies the correct boundary conditions at $t = 0$, differentiate (43) and write η for s. Then

$$\left[\frac{\partial}{\partial t} P(E,t) \right]_{t=0} = \left[\frac{1}{2\pi i E_0} \int_{\sigma-i\infty}^{\sigma+i\infty} d\eta \left\{ \frac{E_0}{E+\beta g(\eta,t)} \right\}^\eta \right.$$

$$\left. \times \left\{ -\frac{\beta \eta f(\eta,t,\beta)}{\{E+\beta g(\eta,t)\}} \frac{\partial g(\eta,t)}{\partial t} + \frac{\partial f(\eta,t,\beta)}{\partial t} \right\} \right]_{t=0}$$

$$= \frac{1}{2\pi i E_0} \int_{\sigma-i\infty}^{\sigma+i\infty} d\eta \left(\frac{E_0}{E} \right)^n \left\{ -\frac{\eta\beta}{E} - A_s \right\}.$$

Multiplying this equation by E^{s-1} and integrating from 0 to ∞, we at once get (39). Hence the function $p(s,t)$ connected with (43) by (13a) rigorously satisfies the correct boundary conditions (21a) and (39), provided $g(s,t)$ and $f(s,t,\beta)$ satisfy (45) and (44) respectively, so that (43) rigorously satisfies the boundary conditions (20) at $t = 0$.

We now proceed to find the equations that $g(s,t)$ and $f(s,t,\beta)$ must satisfy in order that (43) should be an exact solution of (17). Assume for the moment that asymptotically for large s

$$g(s,t) \approx \frac{2}{\alpha' \log s}, \qquad \text{when} \quad \alpha't > 2, \qquad (46a)$$

and

$$g(s,t) \approx t, \qquad \text{when} \quad \alpha't < 2. \qquad (46b)$$

It will be shown in the appendix that $g(s,t)$, as determined by the equations given below, has in fact this asymptotic form for large s. It will also appear that for $R(s)$ greater than some finite positive number, that is in the whole complex plane to the right of some line parallel to the imaginary axis and to the right of it, neither $g(s,t)$ nor $f(s,t,\beta)$ have any singularities.

If $\alpha't > 2$, then by (46a), $g(s,t) \to 0$ as $|s| \to \infty$. Since $g(s,t)$ and $f(s,t,\beta)$ have no singularities for $R(s) > \sigma$, we can always move the contour of the integration in (43) to the right and hence in view of (46) by making σ large enough the path of integration can be changed into another line parallel to the imaginary axis such that

$$E > |\beta g(s,t)| \qquad (47)$$

at every point of it. We can then expand $(E + \beta g)^{-s}$ at every point of this contour as a series in ascending powers of $\beta g(s,t)/E$. If $\alpha't < 2$, then it follows from (46) that a similar contour can be chosen such that on it $(E + \beta g)^{-s}$ can be expanded as a power series in $\beta g(s,t)/E$, provided $E > \beta t$. In order to find $g(s,t)$ and $f(s,t,\beta)$, the expansion may be carried out formally even for $E < \beta t$. (A rigorous proof of the correctness of the solution (43) is given in the paper by Iyengar.) If $f(s,t,\beta)$ satisfies certain conditions at infinity, then the series can be integrated term by term, and

$$P(E,t) = \sum_{n=0}^{\infty} \frac{1}{2\pi i E_0} \int_{\sigma-i\infty}^{\sigma+i\infty} \left(\frac{E_0}{E}\right)^s \frac{\Gamma(s+n)}{\Gamma(s)\,\Gamma(n+1)} \left\{-\frac{\beta g(s,t)}{E}\right\}^n f(s,t,\beta)\,ds. \qquad (48)$$

Writing $\eta = s + n$ in this integral and shifting the path of integration again to the left so that $\sigma + n$ is replaced by σ, the $(n+1)$th term becomes

$$\frac{1}{2\pi i E_0} \left(-\frac{\beta}{E_0}\right)^n \int_{\sigma-i\infty}^{\sigma+i\infty} \left(\frac{E_0}{E}\right)^\eta \frac{(\eta-1)\ldots(\eta-n)}{n!} \{g(\eta-n,t)\}^n f(\eta-n,t,\beta)\,d\eta.$$

Now calculate the function $p(s,t)$ connected with $P(E,t)$ by the transformation $(13a)$. It at once follows from the property of the Mellin transform that this is

$$E_0^{s-1}\left(-\frac{\beta}{E_0}\right)^n \frac{(s-1)\ldots(s-n)}{n!} \{g(s-n,t)\}^n f(s-n,t,\beta).$$

Hence
$$p(s,t) = E_0^{s-1}\left[f(s,t,\beta) - (s-1)\frac{\beta}{E_0} g(s-1,t)f(s-1,t,\beta) \right.$$

$$\left. + \frac{(s-1)(s-2)}{2!}\left\{\frac{\beta g(s-2,t)}{E_0}\right\}^2 f(s-2,t,\beta) + \ldots \right]. \qquad (49)$$

From (49) it can be easily deduced that if $f(s,t,\beta)$ satisfies the boundary conditions (44), then $g(s,t)$ must satisfy (45) in order that $p(s,t)$ should satisfy (21 a) and (39). Conversely if $g(s,t)$ satisfies (45) then $f(s,t,\beta)$ must satisfy (44).

Now assume that $f(s,t,\beta)$ can be expressed as a power series in β thus

$$f(s,t,\beta) = f_0(s,t) + \left(\frac{\beta}{E_0}\right)^2 f_2(s,t) + \left(\frac{\beta}{E_0}\right)^3 f_3(s,t) + \dots, \tag{50}$$

and that the coefficient of β in this series is zero. It will appear presently that this is possible. There are now sufficient conditions to determine g and f uniquely. Substituting (50) in (44) and equating the different powers of β to zero, it follows that $f_0(s,t)$ satisfies the boundary conditions (44), while $f_n(s,t)$ for $n > 0$ satisfies the boundary conditions

$$f_n(s,0) = 0, \quad \left\{\frac{\partial}{\partial t}f_n(s,t)\right\}_{t=0} = 0. \tag{51}$$

Introducing the series (50) into (49) and rearranging the terms to form a power series in ascending powers of β we get

$$p(s,t) = E_0^{s-1}\Bigg[f_0(s,t) - \frac{\beta}{E_0}(s-1)g(s-1,t)f_0(s-1,t)$$

$$+ \left(\frac{\beta}{E_0}\right)^2 \left\{\frac{(s-1)(s-2)}{2!}\{g(s-2,t)\}^2 f_0(s-2,t) + f_2(s,t)\right\} + \dots \Bigg]. \tag{52}$$

Write for brevity $\qquad \varDelta_s \equiv \frac{\partial^2}{\partial t^2} + (A_s + D)\frac{\partial}{\partial t} + (A_s D - B_s C_s). \tag{53}$

Introducing (52) into (17) and equating the coefficients of the different powers of β to zero, we get a set of differential equations which successively determine $g(s,t)$ and $f_n(s,t)$, thus

$$\varDelta_s f_0(s,t) = 0, \tag{54a}$$

$$\varDelta_s\{g(s-1,t)f_0(s-1,t)\} = \left(\frac{\partial}{\partial t}+D\right)f_0(s-1,t), \tag{54b}$$

$$\varDelta_s\left\{f_2(s,t) + \frac{(s-1)(s-2)}{2!}\{g(s-2,t)\}^2 f_0(s-2,t)\right\}$$

$$= \left(\frac{\partial}{\partial t}+D\right)\{(s-1)(s-2)g(s-2,t)f_0(s-2,t)\}, \tag{54c}$$

....

It follows from (54a) and the boundary conditions (44) that $f_0(s,t)$ is precisely the function we had in the case $\beta = 0$, namely,

$$f_0(s,t) = \frac{D-\lambda_s}{\mu_s-\lambda_s}e^{-\lambda_s t} + \frac{\mu_s-D}{\mu_s-\lambda_s}e^{-\mu_s t}. \tag{55}$$

Substituting this into (54b) we get at once

$$g(s-1,t)f_0(s-1,t) = h_s e^{-\lambda_s t} + j_s e^{-\mu_s t}$$
$$+ \frac{(D-\lambda_{s-1})}{(\lambda_{s-1}-\lambda_s)(\lambda_{s-1}-\mu_s)} \frac{(D-\lambda_{s-1})}{(\mu_{s-1}-\lambda_{s-1})} e^{-\lambda_{s-1}t} + \frac{(D-\mu_{s-1})}{(\mu_{s-1}-\lambda_s)(\mu_{s-1}-\mu_s)} \frac{\mu_{s-1}-D}{\mu_{s-1}-\lambda_{s-1}} e^{-\mu_{s-1}t},$$
$$(56)$$

where we have made use of relations of the type

$$\lambda_{s-1}^2 - (A_s+D)\lambda_{s-1} + A_s D - B_s C_s = (\lambda_{s-1}-\lambda_s)(\lambda_{s-1}-\mu_s).$$

λ_s and μ_s are the functions of s defined by (19), and h_s and j_s are arbitrary functions of s independent of t which have to be determined so that g may satisfy the boundary conditions (45). After some calculation we find

$$g(s,t) = \left[\frac{(D-\lambda_s)^2 e^{-\lambda_s t}}{(\mu_s-\lambda_s)(\lambda_{s+1}-\lambda_s)(\mu_{s+1}-\lambda_s)} + \frac{(D-\lambda_{s+1})^2 e^{-\lambda_{s+1}t}}{(\mu_{s+1}-\lambda_{s+1})(\lambda_s-\lambda_{s+1})(\mu_s-\lambda_{s+1})} \right.$$
$$\left. - \frac{(D-\mu_s)^2 e^{-\mu_s t}}{(\mu_s-\lambda_s)(\lambda_{s+1}-\mu_s)(\mu_{s+1}-\mu_s)} - \frac{(D-\mu_{s+1})^2 e^{-\mu_{s+1}t}}{(\mu_{s+1}-\lambda_{s+1})(\lambda_s-\mu_{s+1})(\mu_s-\mu_{s+1})} \right]$$
$$\times \left\{ \frac{D-\lambda_s}{\mu_s-\lambda_s} e^{-\lambda_s t} + \frac{\mu_s-D}{\mu_s-\lambda_s} e^{-\mu_s t} \right\}^{-1}. \qquad (57)$$

Since $\mu_s > \lambda_s$ for real s greater than 1, the terms proportional to $e^{-\mu t}$ are negligible for all except very small t, and then to a good approximation

$$g(s,t) \approx \frac{D-\lambda_s}{(\lambda_{s+1}-\lambda_s)(\mu_{s+1}-\lambda_s)} - \frac{(D-\lambda_{s+1})^2(\mu_s-\lambda_s)e^{-(\lambda_{s+1}-\lambda_s)t}}{(D-\lambda_s)(\mu_{s+1}-\lambda_{s+1})(\mu_s-\lambda_{s+1})(\lambda_{s+1}-\lambda_s)}. \qquad (58)$$

Since, further, $\lambda_{s+1} > \lambda_s$ for real s greater than 1, the second term tends to zero for large t, and $g(s,t)$ becomes independent of t. For all except small t the second term gives a small correction. The values of $g(s,t)$ for $t = 2, 4, 10$ and ∞ as calculated from (58) are given in table 4A. For $t = 2$ and 4 and large s the figures are in error by 10 % as (58) is then not a good enough approximation to (57). This can be seen by comparing these figures with those of table 4B, which are calculated from (57). But just for $t \lesssim 2$ the saddle-point s_0 usually lies near or below 2, and here the differences between the figures of tables 4A and 4B are negligible. The figures for $t = 10$ and ∞ in table 4A are quite accurate. The general dependence of $g(s,t)$ on t for fixed s is easily seen. From the boundary conditions (45) it follows that for small t

$$g(s,t) \sim t, \qquad (59)$$

and it rapidly tends to some constant limiting value depending on s as $t \to \infty$. This limiting value is about $0 \cdot 5$ for $s = 1 \cdot 5$, about $0 \cdot 8$ for $s = 2$, and about 1 for s between $2 \cdot 5$ and $3 \cdot 0$. This behaviour of $g(s,t)$ is of physical significance, as will appear in the next section. From table 4B and (46) it is clear that for large s the curves have a 'dip' near $t \sim 2/\alpha' \approx \frac{2}{3}$, which becomes more pronounced as s increases, until for $s \to \infty$ it becomes a sharp discontinuity at $t = 2/\alpha'$.

TABLE 4A. VALUES OF $g(s, t)$ FROM FORMULA (58)

s	$g(s, 2)$	$g(s, 4)$	$g(s, 10)$	$g(s, \infty)$
1·1	—	—	—	—
1·2	0·0000	—	—	0·3099
1·3	—	—	—	0·3922
1·4	—	—	—	0·4673
1·5	0·5252	0·5353	0·5364	0·5364
1·6	—	—	—	0·6008
1·7	—	—	—	0·6609
1·8	—	—	—	0·7167
1·9	—	—	—	0·7695
2·0	0·7599	0·7981	0·8177	0·8186
2·1	—	—	—	0·8641
2·2	—	—	—	0·9061
2·3	—	—	—	0·9450
2·4	—	—	—	0·9807
2·5	0·895	0·944	0·9993	1·013
2·6	0·913	0·962	—	1·043
2·7	0·927	0·976	—	1·070
2·8	0·939	0·987	—	1·095
2·9	0·950	0·997	—	1·120
3·0	0·958	1·004	1·0835	1·141
3·1	0·962	1·006	1·0882	1·160
3·2	0·967	1·011	1·0946	1·179
3·3	0·972	1·014	1·0984	1·198
3·4	0·972	1·010	1·0944	1·209
3·5	0·971	1·010	1·0944	1·226
3·6	0·968	1·005	1·0895	1·240
3·7	0·966	1·000	1·0835	1·253
3·8	0·965	1·000	1·0817	1·268
3·9	0·965	0·999	1·0782	1·283
4·0	0·958	0·990	1·0675	1·292
4·1	—	—	1·0633	1·304
4·2	—	—	1·0559	1·315
4·3	—	—	1·0453	1·326
4·4	—	—	1·0406	1·339
4·5	—	—	1·0272	1·350
4·6	—	—	1·0198	1·360
4·7	—	—	1·0111	1·375
4·8	—	—	1·0021	1·378
4·9	—	—	0·9968	1·383
5·0	—	—	0·9876	1·386

TABLE 4B. ACCURATE VALUES OF $g(s, t)$ FROM FORMULA (57)

s \ t	1·0	1·3	1·5	2·0	4·0	10·0
1·5	0·4892	0·5113	0·5174	0·5269	0·5353	0·5364
2·0	0·6309	0·7071	0·7155	0·7705	0·8009	0·8186
2·5	0·7264	0·8298	0·8759	0·9339	0·9564	0·9993
3·0	0·7738	0·9035	0·9659	1·048	1·037	1·083
3·5	0·8054	0·9545	1·029	1·129	1·066	1·094
4·0	0·8317	0·9949	1·079	1·189	1·064	1·067
4·5	0·8475	1·023	1·116	1·239	1·047	1·027
5·0	0·8630	1·050	1·149	1·287	1·027	0·9876

If we wish to calculate the limiting form of $g(s, t)$ for large s but finite t, it is necessary to consider all the terms in (57) and (58), for the first and second terms and the third and fourth terms in the numerator of (57) largely compensate each other as $s \to \infty$, as also do the two terms in (58). The limiting form of $g(s, t)$ as $s \to \infty$ is calculated in the appendix and the result has already been given in (46) above.

The expression (57) for $g(s, t)$ like $f_0(s, t)$ given by (55) and indeed, like $f_n(s, t)$ in general, is symmetrical in λ and μ. Hence, although λ and μ are themselves not one-valued functions of s in the whole complex plane, $g(s, t), f_0(s, t)$ and $f_n(s, t)$ are one-valued functions of s throughout. Moreover, since they contain $e^{-\lambda t}$ and $e^{-\mu t}$, it is clear that the points in the s plane where λ and μ have singularities are essential singularities of $g(s, t), f_0(s, t)$ and $f_2(s, t)$. Now λ and μ are singular whenever A_s, B_s or C_s defined by (16) are singular, and this happens when s is equal to 1, 0 or a negative integer.

The functions $f_2(s, t), f_3(s, t)$, etc. can be calculated at once by solving (54c) and the succeeding equations of the set. All these equations are of the same type; in which the right-hand side is known, and the unknown function is on the left-hand side. The general solution of equations of this type is given in the appendix. But (54c) can also be solved directly. We get

$$
f_2(s, t) = -\frac{(s-1)(s-2)}{2} \{g(s-2, t)\}^2 f_0(s-2, t)
$$

$$
+ \frac{\frac{\partial}{\partial t} + D}{\Delta_s} \{(s-1)(s-2) g(s-2, t) f_0(s-2, t)\}
$$

$$
+ k_s e^{-\lambda_s t} + l_s e^{-\mu_s t}. \tag{60}
$$

(56) shows that the term in curly brackets in (60) just contains t in four exponentials, and the operator outside the bracket is to be taken in the sense

$$
\frac{\frac{\partial}{\partial t} + D}{\Delta_s} e^{-\lambda_{s-1} t} = \frac{D - \lambda_{s-1}}{(\lambda_{s-1} - \lambda_s)(\lambda_{s-1} - \mu_s)} e^{-\lambda_{s-1} t}.
$$

The arbitrary functions k_s and l_s which are independent of t, have to be determined to make $f_2(s, t)$ satisfy the boundary conditions (51) at $t = 0$. It is clear from (60) that $f_2(s, t)$ will contain terms proportional to $e^{-\lambda_{s-2} t}$, $e^{-\lambda_{s-1} t}$, $e^{-\lambda_s t}$, and a similar set of three terms with μ instead of λ. Now it will be shown below that for $E_0 \gg \beta$, an appreciable contribution to $P_2(E, t)$ only comes from the term proportional to $e^{-\lambda_{s-2} t}$, except for very small t. Hence it is not necessary to determine k_s and l_s explicitly. The effect of the two terms containing them is largely to cancel the effect of the $e^{-\lambda_{s-2} t}$ term at small t, and thus to make the total contribution of f_2 to the series (50) quite negligible for small t, as will be seen in the next section. The term

proportional to $e^{-\lambda_s - _s t}$ in $f_2(s,t)$ can be calculated quite easily, remembering (56) and (58). We find

$$f_2(s,t) = (s-1)(s-2)\frac{(D-\lambda_{s-2})^3}{(\mu_{s-2}-\lambda_{s-2})(\lambda_{s-1}-\lambda_{s-2})^2(\mu_{s-1}-\lambda_{s-2})^2},$$

$$\times \left\{ \frac{(\lambda_{s-1}-\lambda_{s-2})(\mu_{s-1}-\lambda_{s-2})}{(\lambda_s-\lambda_{s-2})(\mu_s-\lambda_{s-2})} - \frac{1}{2} \right\} e^{-\lambda_{s-2}t}$$

$$+ n_s e^{-\lambda_{s-1}t} + \text{terms containing } e^{-\lambda_s t} \text{ and } e^{-\mu t}. \tag{61}$$

The coefficients of the other terms can be calculated easily.

The functions $f_3(s,t)$, etc., can be calculated explicitly by the method given in the appendix. It is therefore clear that an exact formal solution of equation (17) with the boundary conditions (21 a) and (39) can be written in the form (43) with the function $g(s,t)$ given by (57) and $f(s,t,\beta)$ given by the series (50). Now corresponding to the series (50) for $f(s,t,\beta)$, $P(E,t)$ given by (43) can be written as a series

$$P(E,t) = P_0(E,t) + P_2(E,t) + P_3(E,t) + \dots \tag{62}$$

where

$$P_0(E,t) = \frac{1}{2\pi i E_0} \int_{\sigma-i\infty}^{\sigma+i\infty} \left\{ \frac{E_0}{E+\beta g(s,t)} \right\}^s f_0(s,t)\,ds, \tag{63}$$

$$P_2(E,t) = \frac{1}{2\pi i E_0} \left(\frac{\beta}{E_0}\right)^2 \int_{\sigma-i\infty}^{\sigma+i\infty} \left\{ \frac{E_0}{E+\beta g(s,t)} \right\}^s f_2(s,t)\,ds, \tag{64}$$

and so on. It should be noted that the series (62) is not a simple series in powers of β since, owing to the appearances of β in the denominator of the integrand of each term, each term by itself is a function of β which may be expanded as a power series only in certain circumstances.

We now proceed to calculate (63). The function $f_0(s,t)$ given by (55) is the same as the function in curly brackets of (24). The only difference between (63) and (24) is the appearance of $E+\beta g(s,t)$ in place of E. As in (24) the contribution to (63) which comes from the term proportional to $e^{-\mu_s t}$ in $f_0(s,t)$ is negligible compared with the contribution which comes from the term proportional to $e^{-\lambda_s t}$ for all except very small t. Thus to a very good approximation for all except small t, we may write

$$P_0(E,t) \approx \frac{1}{2\pi i E_0} \int_{\sigma-i\infty}^{\sigma+i\infty} \left\{ \frac{E_0}{E+\beta g(s,t)} \right\}^s \frac{D-\lambda_s}{\mu_s-\lambda_s} e^{-\lambda_s t}\,ds. \tag{65}$$

In view of the properties of $g(s,t)$ discussed above and its asymptotic behaviour for large s, it is clear that (65) can again be evaluated by the saddle-point method and indeed with the same degree of accuracy. The effect of replacing E in (27) by $E+\beta g(s,t)$ in (65) is therefore to give E a larger effective value depending on the position of the saddle-point s_0. The position of the saddle-point s_0 which depends on E and t is shifted to somewhat greater values by the presence of $g(s,t)$, as compared

with the corresponding values it had in (27). In calculating (65) it is convenient to introduce a variable y_0 defined by

$$y_0 = \log \frac{E_0}{\beta}. \tag{66}$$

Writing the integrand of (65) again in the form $e^{\psi(s)}$ we find

$$\psi(s) = y_0 s - \lambda_s t + \log \frac{D - \lambda_s}{\mu_s - \lambda_s} - s \log \left\{ \frac{E}{\beta} + g(s, t) \right\}. \tag{67}$$

This differs from the ψ of (29) only by the addition of $g(s, t)$ in the last term. The rest of the procedure of finding the saddle-point s_0 and then evaluating (65) follows the method of the last section closely. In fact, by using tables 2 and 4 it is no more difficult to evaluate (65) than (27).

The qualitative effect of $\beta g(s, t)$ in the denominator of (63) can be seen at once. It follows from (59) that $E + \beta g(s, t) = E + \beta t$ for small t, so that the effect of this term is to shift the whole spectrum to lower energies by an amount βt corresponding to the energy lost by collision by each particle in the thickness t. For all but very small t, however, $g(s, t) \sim 1$, as shown by table 4A and hence $E + \beta g(s, t) \sim E + \beta$. In other words, instead of the number of particles tending to infinity as $E \to 0$, as it did in (24), the spectrum becomes flat for $E < \beta$ and tends to a finite value as $E \to 0$. Thus for all but small t the effect of collision loss is to flatten out the spectrum for energies below the critical energy. We see that our present calculations justify the original assumption of Bhabha & Heitler concerning the effect of collision loss on the energy spectrum of a shower. The finer features of the energy spectrum will be discussed in the next section.

We now proceed to calculate $P_2(E, t)$ inserting in (64) only the term proportional to $e^{-\lambda_s - 2 t}$ of the expression (61) for $f_2(s, t)$. Writing $s + 2$ in the integrand in place of s and moving the resulting contour again to the right, i.e. bringing it back from $\sigma - 2$ to σ, we get

$$P_2(E, t) = \frac{1}{2\pi i E_0} \int_{\sigma - i\infty}^{\sigma + i\infty} \left\{ \frac{E_0}{E + \beta g(s + 2, t)} \right\}^s \frac{\beta^2}{\{E + \beta g(s + 2, t)\}^2} m_{s+2} e^{-\lambda_s t} ds. \tag{68}$$

m_s denotes the coefficient of $\exp(-d_{s-2} t)$ in (61). E_0 appears here in precisely the same way as in (65) and t also appears in an exponential multiplied by the same coefficient λ_s. (68) now differs from (65) first in having $g(s + 2, t)$ in place of $g(s, t)$. Table 4 shows that for s between 1·5 and 3·0, $g(s + 2, t) \sim g(s, t)$, so that the effect of this difference alone would be to make (68) only slightly different from (65). The other difference between (68) and (65) is that the integrand of the former is

$$\left\{ \frac{\beta}{E + \beta g(s + 2, t)} \right\}^2 \frac{\mu_s - \lambda_s}{D - \lambda_s} m_{s+2} \tag{69}$$

times the integrand of the latter. Even for E as small as β/e and s less than 2·5 this factor is less than $\frac{1}{4}$. This extra factor decreases as s increases and hence always shifts the saddle-point of (68) to somewhat smaller values compared with (65).

But since the saddle-point is precisely the point at which the value of the integrand is stationary, this shift does not greatly affect the value of the integral. Thus, (68) will be smaller than (65) roughly by the factor (69). We now see the great advantage of putting $E + \beta g(s, t)$ instead of E in (43) for it multiplies the nth term of the series (62) by a factor $\beta^n \{E + \beta g(s + n, t)\}^{-n}$, instead of the factor $(\beta/E)^n$ which occurs in (42). This factor is always less than or of the order unity however small E may be, whereas β/E becomes very large for small E, and makes the successive terms in (42) increase rapidly.

It is also clear now why the term proportional to $e^{-\lambda_s - 1^t}$ in (61) would make a contribution to $P_2(E, t)$ small compared with (68). For on introducing this in place of $f_2(s, t)$ in (64), one would have to write $s + 1$ instead of s to bring the exponential to the form $e^{-\lambda_s t}$ as in (68) and this would give an integral like (65) but with the extra factor

$$\frac{\beta}{E_0} \frac{\beta}{E + \beta g(s + 1, t)} \frac{\mu_s - \lambda_s}{D - \lambda_s} n_{s+1}$$

in place of (69). Since $E_0 \gg \beta$ in the cascade theory this is small compared with unity and hence the contribution of the $e^{-\lambda_s - 1^t}$ terms in (61) to $P_2(E, t)$ is quite negligible provided $E_0 \gg \beta$. For smaller E_0 the effect of these terms is actually to compensate (68) and make the total contribution of $P_2(E, t)$ to the series (62) still smaller.

In order to compare the contribution of the second term $P_2(E, t)$ of the series (62) compared with the first $P_0(E, t)$ we have worked out (68) accurately by the saddle-point method for $t = 10$, and y_0 ranging from 4 to 10 at intervals of 1, and for the two cases where E is equal to β and β/e respectively. The results are given in table 5. The first and second rows in each case give the values of $\beta P_0(E, t)$ and $\beta P_2(E, t)$ with the corresponding positions of the saddle-point. The third row gives the corresponding figures for $\beta P_\lambda(E, t)$ given by (27) of the previous section, in which collision loss is neglected completely. The table shows clearly that the contribution of P_2 is always considerably smaller than P_0, and for large y_0 it is about one-fifth. Now the thickness $t = 10$ is approximately that at which a particle of initial energy corresponding to $y_0 \sim 12$ produces the greatest number of particles, while for $y_0 = 4$ it corresponds to a position far beyond the maximum, when the shower is getting absorbed by collision loss. We should therefore expect P_2 to give a greater relative contribution for the smaller values of y_0 in table 5, as indeed the figures show. Even so, the table shows that the figures calculated by using the first term $P_0(E, t)$ alone of the series (62) will not differ from the true figures by more than about 30 % except when $E \sim \beta/e$. On the other hand, the figures for the number of particles calculated by neglecting collision loss altogether are three to five times too large for electrons of the critical energy ($E = \beta$), and eight to seventeen times too large for $E = \beta/e$.

When $E \ll \beta/e$, the factor (69) becomes practically equal to unity and then $P_2(E, t)$ becomes nearly as great as $P_0(E, t)$. We should then expect that the higher terms of the series (62) would make an appreciable contribution and that it would be quite insufficient to stop at the first term. The range of energy concerned is small com-

pared with the critical energy, so that the region is not important as far as the total number of particles is concerned.

TABLE 5. COMPARISON OF $P_0(E, t)$ AND $P_2(E, t)$ FOR $t = 10$

	y_0	4	5	6	7	8	9	10
	s_0	3·19	2·86	2·63	2·46	2·32	2·21	2·12
	$\beta P_0(E, t)$	0·0819	0·456	2·155	8·822	32·37	100·5	306·0
$E = \beta$	s_0	2·78	2·53	2·36	2·24	2·15	2·06	2·00
	$\beta P_2(E, t)$	0·0374	0·1614	0·585	1·944	5·842	15·00	40·72
	$\beta P_\lambda(E, t)$	0·591	2·84	11·4	40·7	130·0	377·0	1012
	s_0	3·10	2·80	2·59	2·43	2·30	2·19	2·10
	$\beta P_0(E, t)$	0·2560	1·209	5·472	21·90	78·97	247·2	715·8
$E = \beta/e$	s_0	2·66	2·45	2·30	2·19	2·11	2·02	1·95
	$\beta P_2(E, t)$	0·1994	0·7820	2·789	8·782	24·99	65·17	157·5
	$\beta P_\lambda(E, t)$	7·72	30·9	110	354	1024	2751	6912

Since the second term P_2 of the series (62) is of the order β^2, it might appear that stopping at the first term means a total neglect of terms of the order β^2. This is, however, not so, since the integrand of the first term given by (63) itself contains β in an expression which can be expanded as a power series in β, as has been done in (48). To compare the contributions to the term of order β^2 in $P(E, t)$ which come from P_0 and P_2 respectively, calculate $P(E, t)$ by using the transformation (14a) on the series (52). We get*

$$P(E, t) = \frac{1}{2\pi i E_0} \int_{\sigma - i\infty}^{\sigma + i\infty} \left(\frac{E_0}{E}\right)^s \left[f_0(s, t) - \frac{\beta}{E_0} (s - 1) g(s - 1, t) f_0(s - 1, t) \right.$$

$$\left. + \left(\frac{\beta}{E_0}\right)^2 \left\{ \frac{(s - 1)(s - 2)}{2} \{g(s - 2, t)\}^2 f_0(s - 2, t) + f_2(s, t) \right\} + \dots \right]. \quad (70)$$

In this series the term independent of β, and the term of order β are entirely due to the first term $P_0(E, t)$. Of the contributions to the term of order β^2 the first part in curly brackets in (70) is due to $P_0(E, t)$ while the second is due to $P_2(E, t)$. To get an idea of the relative magnitude of the two terms in curly brackets in (70) consider that part of both which has $e^{-\lambda_s - s t}$ as a factor. Using (58), (57) and (61), and writing s for $s - 2$ for convenience, this part becomes

$$\left[1 + \left\{ 2 \frac{(\lambda_{s+1} - \lambda_s)(\mu_{s+1} - \mu_s)}{(\lambda_{s+2} - \lambda_s)(\mu_{s+2} - \mu_s)} - 1 \right\} \right],$$

multiplied by a function of s which we have not written explicitly for brevity. The term in curly brackets is the contribution of $f_2(s, t)$. The expression in curly brackets

---

* The same result could be obtained after some elementary manipulation by introducing the series (50) into (48).

is 0·52 for $s = 1·5$, 0·40 for $s = 2$ and 0·32 for $s = 2·5$. Using (25) it can be shown that it tends to zero as $s \to \infty$. It is clear from this that *at least* two-thirds of the term proportional to β^2 in the strict expansion of $P(E, t)$ as a power series in β comes from P_0, while P_2 only contributes less than 35 % to this term. We see therefore that the first term P_0 of the series (62) already contains the whole contributions to $P(E, t)$ of the terms independent of β, and proportional to β, and the major part of the contribution of the term proportional to β^2 and probably the important part of the contribution of higher powers of β. The terms P_2, P_3, etc., then only give small corrections to the terms proportional to β^2 and higher powers of β.

We have therefore established that (62) is an accurate formal solution of the equations of the cascade theory taking collision loss into account. The convergence of the series (62) has not been proved, but this is not necessary since we have shown that *in general it is sufficient to take only the first term*, the second being small compared with the first, *even for energies below the critical energy*, so that it is possible to look upon (62) as an asymptotic solution of the problem should the series (62) not converge.*

Values of $P_0(E, t)$ have been calculated by the saddle-point method for y_0 between 3 and 10 at intervals of 1, and three values of E equal to $e\beta$, β and β/e respectively. Table 6 gives the figures for $t = 2$, table 7 for $t = 4$, and table 8 for $t = 10$. The corresponding values of s_0 are given in table 7 as an indication. The figures for $P_\lambda(E, t)$ are those given by formula (27) where collision loss is neglected completely. *The tables clearly show that for large t and small y_0 the figures given hitherto, neglecting collision loss altogether, are too large by a factor two even when E is e times the critical energy, and for E equal to the critical energy, the old figures neglecting collision loss are sometimes too large by a factor seven.* This is as we should expect, for it is just for relatively large t and small y_0, that is after the shower has passed its maximum, that collision loss has the greatest effect. In this region our calculations show that the previous theories in which collision loss was neglected completely cannot claim accuracy even for energies which are two or three times the critical energy.

ENERGY DISTRIBUTION OF THE CASCADE ELECTRONS

Small thicknesses. First consider the energy distribution of cascade electrons in a shower as it appears after passing through a very thin layer of substance such that $t \ll 1$. (62) is still a correct solution, but in $f_0(s, t)$, $f_2(s, t)$, etc., it is no longer possible to neglect the terms containing $e^{-\mu t}$. For $t \ll 1$ it is therefore simpler to start directly from the expression (43) and to insert in it for g and f the expressions they assume for very small t. Since f_0 satisfies the boundary conditions (44) it follows from (54a) that

$$\left\{ \frac{\partial^2}{\partial t^2} f_0(s, t) \right\}_{t=0} = A_s^2 + B_s C_s.$$

* The convergence has been proved in the paper by Iyengar (1942).

TABLE 6. $t = 2$

	y_0	3	4	5	6	7	8	9	10
$E = e\beta$	$\beta P_0(E,t)$	0·259	0·503	0·753	1·12	1·62	2·18	2·86	3·84
	$\beta P_\lambda(E,t)$	0·391	0·619	0·955	1·39	1·91	2·59	3·45	4·53
$E = \beta$	$\beta P_0(E,t)$	0·831	1·519	2·27	3·34	4·56	6·19	8·16	10·7
	$\beta P_\lambda(E,t)$	1·68	2·60	3·78	5·19	7·05	9·37	12·3	15·3
$E = \beta/e$	$\beta P_0(E,t)$	1·87	3·14	4·97	7·14	9·89	13·6	17·9	23·5
	$\beta P_\lambda(E,t)$	7·05	10·3	14·1	19·2	25·5	33·5	41·6	49·7

TABLE 7. $t = 4$

	y_0	3	4	5	6	7	8	9	10
$E = \beta e$	s_0	—	2·63	2·24	2·01	1·86	1·75	1·67	1·61
	$\beta P_0(E,t)$	—	0·333	0·989	2·39	4·90	9·41	16·6	28·3
	s_0	2·96	2·39	2·10	1·92	1·79	1·71	1·64	1·58
	$\beta P_\lambda(E,t)$	0·171	0·609	1·60	3·52	6·88	13·0	21·1	36·0
$E = \beta$	s_0	3·00	2·43	2·13	1·94	1·81	1·72	1·64	1·59
	$\beta P_0(E,t)$	0·385	1·29	3·61	8·00	16·3	30·3	52·0	87·9
	s_0	2·39	2·10	1·92	1·79	1·71	1·64	1·58	1·54
	$\beta P_\lambda(E,t)$	1·65	4·34	9·57	18·7	35·2	57·3	97·7	160·0
$E = \beta/e$	s_0	2·84	2·37	2·09	1·92	1·79	1·71	1·64	1·58
	$\beta P_0(E,t)$	0·931	3·11	8·13	18·5	35·8	67·4	116·0	189·0
	s_0	2·10	1·92	1·79	1·71	1·64	1·58	1·54	1·50
	$\beta P_\lambda(E,t)$	11·8	26·0	50·8	95·8	155·8	266·0	435·0	671·0

TABLE 8. $t = 10$

	y_0	4	5	6	7	8	9	10
$E = \beta e$	$\beta P_0(E,t)$	0·0138	0·0870	0·471	2·04	7·89	26·1	80·6
	$\beta P_\lambda(E,t)$	0·0379	0·217	1·05	4·18	15·0	47·9	139·0
$E = \beta$	$\beta P_0(E,t)$	0·0819	0·456	2·15	8·82	32·4	100·0	306·0
	$\beta P_\lambda(E,t)$	0·591	2·84	11·4	40·7	130·0	377·0	1012
$E = \beta/e$	$\beta P_0(E,t)$	0·256	1·21	5·47	21·9	79·0	247·0	716·0
	$\beta P_\lambda(E,t)$	7·72	30·9	110·0	354·0	1024	2751	6912

This together with (44) shows that for small t

$$f_0(s,t) \approx 1 - A_s t + \tfrac{1}{2}(A_s^2 + B_s C_s)t^2 + O(t^3) \qquad (71)$$

as could be derived after some calculation from (55). Further, in view of the boundary conditions (44) and (45), it follows from (54b) that

$$\left\{\frac{\partial^2}{\partial t^2}g(s,t)\right\}_{t=0} = A_s - A_{s+1}.$$

Hence for small t $\qquad g(s,t) \approx t - \tfrac{1}{2}(A_{s+1} - A_s)t^2 + O(t^3).$ $\qquad (72)$

By using the boundary conditions (44), (45) and (51) it can then be deduced from (54c) and the subsequent equations that $\partial^2/\partial t^2 f_n(s,t) = 0$ for all $n > 0$, so that $f_n(s,t)$ is of order t^3. All the following terms in (50) are therefore negligible compared with

f_0 for small t. It follows from this that for small t all the terms in (62) except the first are of order t^3 and negligible compared with $P_0(E, t)$. Hence, inserting (71) and (72) into (43),

$$P(E, t) = \frac{1}{2\pi i E_0} \int_{\sigma-i\infty}^{\sigma+i\infty} \left\{ \frac{E_0}{E + \beta[t - \frac{1}{2}(A_{s+1} - A_s) t^2]} \right\}^s [1 - A_s t + \frac{1}{2}(A_s^2 + B_s C_s) t^2] ds.$$

Correct to terms of order t^2 this may equally well be written after some elementary transformations

$$P(E, t) = \frac{1}{2\pi i E_0} \int_{\sigma-i\infty}^{\sigma+i\infty} \left(\frac{E_0}{E + \beta t} \right)^s \left[1 - A_s t + \frac{1}{2}(A_s^2 + B_s C_s) t^2 + \frac{\beta}{E_0} \frac{s-1}{2} (A_s - A_{s-1}) t^2 \right] ds.$$
(73)

After inserting the expressions (16) for A_s, B_s and C_s in (73) the resulting integral can be evaluated exactly. We give only the result here. It is convenient to introduce a quantity y' defined by

$$y' = \log \frac{E_0}{E + \beta t}.$$
(74)

The 1 in the square brackets in (73) gives as usual

$$\delta(E_0 - E - \beta t).$$
(75)

This, of course, just represents the original electron which has lost an energy βt by collision.

The term proportional to t in the square brackets in (73) only contains A_s. It is obvious from the way that equation (15a) was derived from (12a) that A_s represents the effects of radiation loss only, so that the spectrum of electrons created in a very small thickness is due to one radiation process only. The term proportional to t in the square brackets in (73) gives

$$-(\alpha'\kappa - \alpha' + \frac{1}{2}) t\delta(E_0 - E - \beta t) + \frac{t}{E_0} \left(\frac{\alpha'}{e^{y'} - 1} + 1 - e^{-y'} \right),$$
(76)

where $$\kappa = \int_0^1 \frac{dZ}{Z}.$$

κ is therefore an infinite constant which multiplies the δ function. The spectrum reaches a finite value as $E \to 0$, increases with increasing E, and tends to infinity as $E \to E_0 - \beta t$, that is as $y' \to 0$. This form of the spectrum is quite understandable, for the spectrum of low energy quanta given by formula (1) is of the form dE'/E', so that the probability is greatest for the electron losing a very small part of its energy. The δ function in (76) diminishes the δ function in (75) corresponding to the removal of the initial electron by radiation loss. The diminution is, however, infinite, just because the total chance of a quantum being emitted is infinite according to (1). Both the singularities in (76) are due entirely to a failure of the Bethe-Heitler formula (1) for quanta of very low energy, and here, as has been shown by Bloch & Nordsieck

(1937), we have to consider the simultaneous emission of a large number of quanta involved in a transition to the classical theory. (76) could of course be derived by a direct integration of (12a). The resulting electron spectrum is given in the third column of table 9 for $t = 0\cdot1$.

TABLE 9. ENERGY SPECTRUM OF ELECTRONS AT DIFFERENT THICKNESSES FOR
$$y_0 \equiv \log E_0/\beta = 5. \quad y = \log (E/\beta)$$

E/β	y	(76)	(77)	(78)	Total				
		t 0·1	0·1	0·1	0·1	0·3	2·0	4·0	10·0
0·37	−1	0·1001	0·0274	3·259	3·387	20·94	736·95 (2094)	1206·0 (7539)	179·4 (4582)
1	0	0·1003	0·0231	1·356	1·479	11·74	337·3 (560·3)	535·4 (1420)	67·67 (421·7)
2·72	1	0·1007	0·0185	0·5035	0·6227	4·669	111·7 (141·7)	146·8 (237·0)	12·91 (32·20)
7·39	2	0·1022	0·0137	0·1709	0·2868	1·920	33·83	33·24	2·063
20·1	3	0·1078	0·0090	0·0512	0·1680	0·8583	7·865	3·443	——
54·7	4	0·1424	0·0042	0·0121	0·1587	0·5739	——	——	——
——	4·8	0·6315	−0·0862	0·0019	0·5472	1·139	——	——	——

The figures in brackets are those neglecting collision loss. All the other figures represent $E_0 P_0(E, t)$.

The first term proportional t^2 in the square brackets in (73) has a part proportional to A_s^2, representing the spectrum created by two successive radiation processes, which merely results in a spectrum of the general form (76) but still more flattened out. It gives

$$(\alpha'\kappa - \alpha' + \tfrac{1}{2})^2 t^2 \delta(E_0 - E - \beta t) + \frac{t^2}{E_0}\left[\alpha'\left(\frac{\alpha'}{e^{\nu'} - 1} + 1 - e^{-\nu'}\right)\log(1 - e^{-\nu'})\right.$$
$$\left. + \frac{1}{2}\left\{\frac{\alpha'^2}{e^{\nu'} - 1} + 1 + (1 - 2\alpha')e^{-\nu'}\right\}y' + \frac{\alpha'(\alpha' - \tfrac{1}{2})}{e^{\nu'} - 1} + (\alpha' - \tfrac{3}{2})(1 - e^{-\nu'})\right]. \quad (77)$$

The resulting spectrum is given in the fourth column of table 9 for $t = 0\cdot1$. The other part is proportional to $B_s C_s$. It is obvious from the way these quantities were derived that C_s represents the creation of quanta by radiation, while B_s represents the creation of pairs by quanta, so that this part represents the electron spectrum produced by the original electron through the intermediary of one quantum. The resulting spectrum due to this part is

$$\frac{t^2}{E_0}\left[\alpha'D(e^{\nu'} - e^{-2\nu'}) + \left(1 - \alpha' + \frac{\alpha'^2}{2}\right)(1 - e^{-\nu'}) - \alpha'y'(1 + e^{-\nu'})\right]. \quad (78)$$

It tends to zero as $y' \to 0$, i.e. $E \to E_0 - \beta t$, and increases monotonically as E decreases, reaching its maximum value when $E \to 0$. This is again what we should expect, for the spectrum of low energy quanta being of the form dE'/E', the spectrum of electrons created by these quanta rises to a maximum as $E \to 0$. It should be noticed that owing to the appearance of $e^{\nu'}$ in (78) its order of magnitude is in general much

larger than (76) or (77). The spectrum due to (78) is given by the fifth column of table 4 for $t = 0.1$.

The last term in square brackets in (73) is proportional to β. It can be written

$$-\frac{\beta t}{2}\frac{\partial}{\partial E}\left[(1-e^{-\nu'})\left\{-\frac{t}{2\pi i E_0}\int_{\sigma-i\infty}^{\sigma+i\infty}e^{\nu's}A_s\,ds\right\}\right] = \frac{\beta t^2}{E_0^2}\left(1-\frac{\alpha'}{2}-e^{-\nu'}\right). \quad (79)$$

The expression in curly brackets is just equal to (76) and is the spectrum created by one radiation process only. The meaning of the term (79) is then the following. In (76) the spectrum proportional to t has been shifted down to smaller energies by an amount βt, as if all the electrons had lost this amount of energy by collision. This is obviously incorrect, since some of the electrons are created at some intermediate point of the layer and hence lose less energy than βt by collisions. The purpose of the expression (79) is to correct for this error. Moreover, the energy of the primary particle is not E_0, as it appears in (76), at every point of the layer, but less than this by the amount lost by collision. This also introduces a slight correction to the spectrum (76), and this correction is included in (79). It should be noticed that since $E_0 > \beta t$ always, (79) is of order t, and is hence small compared with 1 for small t. It is always small compared with the sum of (76), (77) and (78). In the fifth and sixth columns of table 9 we have given E_0 times the resulting spectrum, namely, E_0 times the sum of (76), (77), (78) and (79) for $t = 0.1$ and $t = 0.3$, and $y_0 = 5$. The spectrum is quite different from what has been generally supposed. There is, as has been shown above, a δ-function at $E = E_0 - \beta t$, the height of which decreases as t increases. A more detailed investigation by K. S. K. Iyengar in which terms of all orders in t have been considered shows that for $t > 1/\alpha'$ this δ-function at $E = E_0 - \beta t$ completely disappears and the spectrum then decreases monotonically as E increases from zero. In particular it shows that the spectrum given by Arley for small thicknesses is not even qualitatively correct. This is mainly due to his neglect of the electron spectrum produced through the intermediary of one quantum. But quite apart from this, his assumptions about the collision and radiation loss are too artificial to have any claim to physical reality.

Formulae (75) to (79) could be derived for small t from a direct integration of the equations (12). Thus every part of the expression (73) has a direct physical meaning and this affords another verification that the solution (43) is correct for small t. In deriving expression (73) from (43), no assumption was made about the magnitude of E_0. Indeed, (43) is also valid when $E_0 < \beta$. As mentioned in the first section, collision loss alone would prevent the electron from penetrating to a thickness t greater than E_0/β, which in this case is small compared with 1. Hence, for $E_0 \ll \beta$, (43) automatically reduces to (73), and this is therefore the complete solution of the problem of the absorption of a low-energy electron by collision loss and cascade production.

Large thicknesses. For large thicknesses the solution (43) can be calculated in the form (62), where as was shown in the last section, the major part of the contribution comes from the first term alone, and it is possible in general to restrict oneself to this term. The form of the spectrum can be obtained from tables 6–8. In table 9 the

spectra for $t = 2, 4$ and 10 have again been given for $y_0 = 5$. The figures in brackets give the spectra that would have been obtained if collision loss had been neglected. Even for energies near the critical energy, the figures neglecting collision loss are sometimes six times too large. It is clear from the correct figures that the effect of collision loss is to flatten the spectrum for $E < \beta$ provided t is not too small. The spectrum of electrons in a shower never falls off for $E < \beta$ as has been suggested by Arley.

The calculation of the total number of particles in a cascade as a function of E_0 and t has been carried out by us on the basis of (62) in another paper (1942).

APPENDIX

Consider the function

$$X(s, t) = \int_0^t f_0(s, t - t') \, Y(s, t') \, dt', \tag{80}$$

where $f_0(s, t)$ is the function defined by (55) which satisfies the equation (54a) and the boundary conditions (44) at $t = 0$. Then

$$\frac{\partial}{\partial t} X(s, t) = Y(s, t) + \int_0^t \frac{\partial}{\partial t} f_0(s, t - t') \, Y(s, t') \, dt',$$

$$\frac{\partial^2}{\partial t^2} X(s, t) = \frac{\partial}{\partial t} Y(s, t) - A_s \, Y(s, t) + \int_0^t \frac{\partial^2}{\partial t^2} f_0(s, t - t') \, Y(s, t') \, dt'.$$

Hence, remembering that f_0 satisfies (45a),

$$\Delta_s X(s, t) = \left(\frac{\partial}{\partial t} + D \right) Y(s, t). \tag{81}$$

Hence, if a function $X(s, t)$ is to be found satisfying an equation of the type (81) in which the right-hand side is given, the particular integral is at once given by (80). To it may be added the complementary function satisfying the left-hand side of (81) equated to zero, in order that X may satisfy the given boundary conditions at $t = 0$.

Now all the equations (54) are of the form (81). Hence, from (54b),

$$g(s - 1, t) f_0(s - 1, t) = \int_0^t f_0(s, t - t') f_0(s - 1, t') \, dt'. \tag{82}$$

This already satisfies the correct boundary conditions (45) at $t = 0$. It leads at once to (57). The subsequent equations may be solved in the same way, where the direct method given in the text fails for $n > 2$.

To calculate the asymptotic forms of $f_0(s, t)$ and $g(s, t)$ for large s we notice that according to (25), as $s \to \infty$,

$$D - \lambda_s \to \frac{2}{\alpha' s^2 \log s}, \tag{83a}$$

while

$$\lambda_{s+1} - \lambda_s \to \frac{4}{\alpha' s^3 \log s}. \tag{83b}$$

Hence, together with (25), for large s

$$f_0(s,t) \to \frac{2}{\alpha'^2 s^2 (\log s)^2} e^{-Dt} + \frac{1}{s^{\alpha't}} e^{-(\alpha'\gamma - \alpha' + \frac{1}{2})t}.$$ (84)

f_0 therefore tends to zero quite differently depending on whether $\alpha't$ is less than or greater than 2, for in the former case the second term, in the latter the first term, of (84) becomes the dominant one as $s \to \infty$.

Now calculate the asymptotic form of $g(s,t)$ given by (57). The first two terms in the numerator must be taken together, and in view of (83b) writing

$$e^{-(\lambda_{s+1} - \lambda_s)t} \approx 1 - (\lambda_{s+1} - \lambda_s)\,t,$$

these then give

$$\frac{e^{-\lambda_s t}}{\lambda_{s+1} - \lambda_s} \left[\frac{(D-\lambda_s)^2}{(\mu_s - \lambda_s)(\mu_{s+1} - \lambda_s)} - \frac{(D-\lambda_{s+1})^2}{(\mu_{s+1} - \lambda_{s+1})(\mu_s - \lambda_{s+1})} + \frac{(D-\lambda_{s+1})^2(\lambda_{s+1} - \lambda_s)\,t}{(\mu_s - \lambda_{s+1})(\mu_{s+1} - \lambda_{s+1})} \right].$$

The last term in the above square brackets is of higher order, and the first two give, neglecting terms of a higher order,

$$e^{-Dt}\left[\frac{2D - \lambda_s - \lambda_{s+1}}{(\mu_s - \lambda_s)(\mu_{s+1} - \lambda_s)} \right] \approx e^{-Dt}\left[\frac{4}{\alpha'^3 s^2 (\log s)^3} \right].$$ (85)

The third and fourth terms in the numerator of (57) must also be taken together. In view of (25)

$$\mu_{s+1} - \mu_s \to \frac{\alpha'}{s}$$

for large s, so that writing

$$e^{-(\mu_{s+1} - \mu_s)t} \to 1 - (\mu_{s+1} - \mu_s)\,t,$$

the third and fourth terms of the numerator of (60) then give

$$\frac{e^{-\mu_s t}}{\mu_{s+1} - \mu_s}\left[\frac{(\mu_s - D)^2}{(\mu_s - \lambda_s)(\mu_s - \lambda_{s+1})} - \frac{(\mu_{s+1} - D)^2}{(\mu_{s+1} - \lambda_s)(\mu_{s+1} - \lambda_{s+1})} + \frac{(\mu_{s+1} - D)^2(\mu_{s+1} - \mu_s)\,t}{(\mu_{s+1} - \lambda_s)(\mu_{s+1} - \lambda_{s+1})} \right].$$

The last term is of order $(\mu_{s+1} - \mu_s)\,t$, while the first two combine together to give a term of higher order. Hence, as $s \to \infty$ the above expression reduces to

$$t e^{-\mu_s t} \to \frac{t}{s^{\alpha't}} e^{-(\alpha'(\gamma - 1) + \frac{1}{2})t}.$$

Adding this to (85) and dividing by (84) the asymptotic form of $g(s,t)$ for large s is

$$g(s,t) \to \frac{\dfrac{4}{\alpha'^3 s^2 (\log s)^3} e^{-Dt} + \dfrac{t}{s^{\alpha't}} e^{-(\alpha'(\gamma-1)+\frac{1}{2})t}}{\dfrac{2}{\alpha'^2 s^2 (\log s)^2} e^{-Dt} + \dfrac{1}{s^{\alpha't}} e^{-(\alpha'(\gamma-1)+\frac{1}{2})t}}.$$ (86)

Depending on whether $\alpha't$ is greater or less than 2, the first or the second terms in the numerator and denominator of (86) dominate, leading to (46a) and (46b) respectively.

REFERENCES

Arley 1938 *Proc. Roy. Soc.* A, **168**, 519–45.
Bethe & Heitler 1934 *Proc. Roy. Soc.* A, **146**, 83–112.
Bhabha & Chakrabarty 1942 *Proc. Indian Acad. Sci.* A, **15**, 464–476.
Bhabha & Heitler 1937 *Proc. Roy. Soc.* A, **159**, 432–58.
Bloch & Nordsieck 1937 *Phys. Rev.* **52**, 54–9.
Carlson & Oppenheimer 1937 *Phys. Rev.* **51**, 220–31.
Chakrabarty 1942 *Proc. Nat. Inst. Sci. India*, **8** (in press).
Corben 1941 *Phys. Rev.* **60**, 435–9.
Iyengar 1942 *Proc. Indian Acad. Sci.* A, **15**, 195–229.
Landau & Rumer 1938 *Proc. Roy. Soc.* A, **166**, 213–28.
Serber 1938 *Phys. Rev.* **54**, 317–20.
Snyder 1938 *Phys. Rev.* **53**, 960–5.
Whittaker & Watson 1927 *Modern analysis*, 4th ed. Camb. Univ. Press.

A magnetic study of the two-phase iron-nickel alloys.   II

By K. Hoselitz and W. Sucksmith, F.R.S.

H. H. Wills Physical Laboratory, University of Bristol

The method of using measurements of magnetic saturation intensity of annealed iron-nickel alloys for the determination of the equilibrium phase boundaries, as demonstrated by Pickles & Sucksmith, has been extended. The phase diagram of the system has been determined accurately between 525 and 365° C. The mechanism of phase segregation from the single-phase α-state has been studied, where it was found that contrary to the usual case, one of the phases crystallizes out in its equilibrium concentration whilst the residue of the alloy progressively and uniformly approaches equilibrium composition. It was possible to study and express quantitatively the rate of attainment of equilibrium, and on evidence obtained in this way the view is based that the lower practical limit of temperature where the equilibrium diagram can be studied by annealing experiments has been reached.

INTRODUCTION

In a recent paper Pickles & Sucksmith (1940) described an investigation of the magnetic properties of the two-phase iron-nickel alloys. Magnetic measurements showed the existence of a two-phase field and the phase diagram above 450° C was determined. The relations between the magnetization temperature curves and phase changes suggested that the magnetic method of investigating the phase diagram was capable of being extended to lower temperatures, if a method could be evolved which allowed the extrapolation to equilibrium conditions from inter-mediate stages of phase segregation. It was therefore the object of the present work to study the mechanism and kinetics of phase changes in the iron-nickel system and to extend the phase diagram so far as possible to lower temperatures.

Pure iron-nickel alloys ranging from 3 to 50 atomic % nickel were obtained from Dr Bradley, their purity and preparation being described before (Bradley & Goldschmidt 1939). Measurements of σ, the magnetic saturation intensity per unit mass, were carried out at various temperatures. High magnetic fields ($\sim$16,000 gauss) were used since the values of σ so observed are independent of physical properties and depend only upon the phases* of the ferromagnetic material. As will be seen, curves of σ as a function of the temperature lend themselves readily to an exhaustive examination of the phases present in a ferromagnetic specimen. The experimental determination of the intensity-temperature variations was the same as used before (Pickles & Sucksmith 1940) and a detailed description of the apparatus and method has been given in an earlier paper (Sucksmith 1939).

As a preliminary heat treatment the specimens were kept at 1000° C for 4 hr. and subsequently quenched in water. They were then cooled to $-180°$ C for 48 hr. The rapid quenching ensured that all the alloys were left in a pure single-phase state, those below 33 % Ni coming into the body-centred cubic α-state, only the Ni richer alloys remaining in the face-centred γ-state. This quenched state formed the most suitable basis for further annealing experiments, because it left the greatest possible part of the range of alloys in the α-state, and only the α-lattice shows perceptible changes after reasonable times of anneal at the low temperatures employed in this work. Specimens left in the γ-state show no change towards equilibrium, even after prolonged times at these temperatures. It was found that the rate of quenching from 1000° C influences the resulting α-lattice and further that the previous method of air cooling did not yield sufficiently consistent results owing to the dependence of the rate of cooling on the size of the specimen and heat capacity of the quartz tubes in which the specimens were sealed. It seems reasonable to assume that this is due to the fact that specimens so cooled exhibit some separation into two phases whilst passing through the two-phase region, when diffusion is still comparatively easy, so that a good single-phase specimen can only be obtained by as rapid quenching as possible.

Heat treatments were carried out at various temperatures down to 325° C for suitable periods. For this purpose large thermostatically controlled electric resistance furnaces were used enabling the temperature to be maintained constant to within $\pm 2°$ C for periods of the order of months. For reasons explained later, most of the alloys were subsequently cooled once more to $-180°$ C for 12 hr.

THE EQUILIBRIUM DIAGRAM

Pickles & Sucksmith (1940), using the magnetic method, showed that it was possible to detect the presence of two phases in some alloys after annealing. They

---

* It should be noted, however, that a superlattice may have a different intensity from that of the same disordered alloy. Information bearing on this will be published in another place.

determined the phase boundaries by comparing the magnetic properties of quenched and annealed alloys. In that case the method of fixing the α-boundary consisted in observing a change in magnetic saturation intensity at room temperature, any γ-phase formed out of the original single-phase α-lattice being evident by a decrease in intensity. It is pointed out by Pickles & Sucksmith that for the lower annealing temperatures this method is not so suitable, as such changes are very small. Furthermore the room temperature intensity even in the highest fields employed depends also to a certain extent on the rate of cooling, and generally on the state of purity of the phase. In the present work another method has been devised which is less sensitive to the conditions of quenching and does not depend on the observation of minute changes in intensity. It has the additional advantage of giving the α- and γ-phase boundaries simultaneously, even if the alloys have not reached complete equilibrium in the annealing treatment. This method is described below.
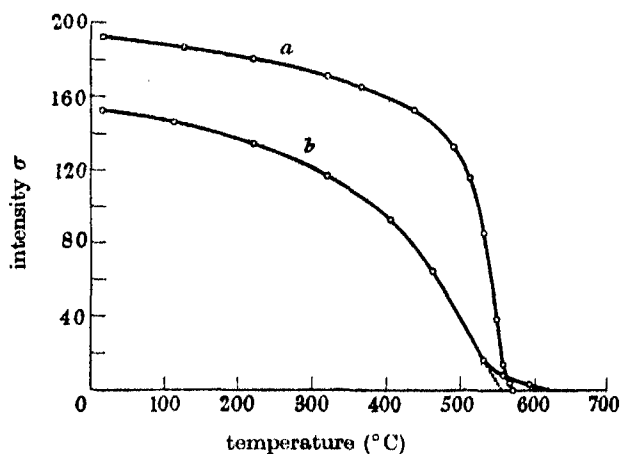
FIGURE 1. Intensity temperature curves of quenched alloys. (*a*) 23% Ni alloy, characteristic α-phase, (*b*) 50% alloy, characteristic γ-phase.

It is known from previous work (Pickles & Sucksmith 1940) that a definite relation exists for each of the two lattice forms between the transition temperature and the nickel content of an alloy. Figure 1 shows two typical $\sigma - T$ curves for single-phase quenched alloys. Curve (*a*) was obtained for a 23% Ni alloy; this has an α-lattice and the graph shows the characteristic features of comparatively constant intensity at lower temperatures with a subsequent sharp fall towards the well-defined transition temperature at 560° C. The non-reversibility of this transition temperature (Pickles & Sucksmith 1940) shows that an irreversible change into the γ-phase has taken place. Curve (*b*), corresponding to a 50% Ni alloy, illustrates the characteristics of a γ-phase by the gradual fall of intensity with rising temperature and the ill-defined Curie point at about 560° C. This transformation is purely magnetic and hence is reversible, the lattice retaining its structure throughout. The point is therefore a true Curie point. Such $\sigma - T$ curves

of well-quenched alloys are repeatable to a high degree of coincidence. Figure 2 is obtained by plotting the transition temperatures in the case of alloys in the α-phase, and the Curie points for γ-alloys, determined from the $\sigma - T$ curves of pure single-phase quenched alloys against their nickel content.

Using figure 2 it is therefore possible to determine the percentage nickel of an unknown single-phase alloy from its $\sigma - T$ curve. This relation allows the determination of the composition in the case of an α-lattice to within $\pm 1 \%$ Ni and in the case of an alloy in the γ-state, where the transition is not so sharp, to within ± 1 to 3% Ni.
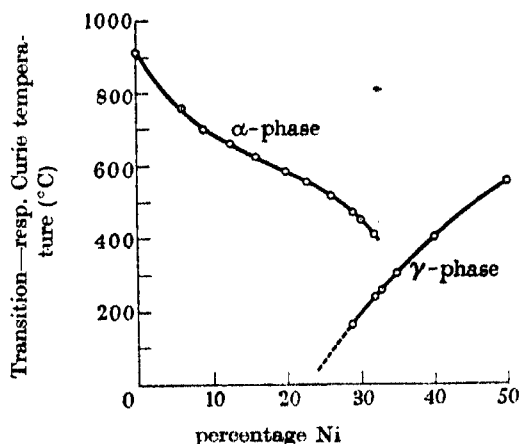
FIGURE 2. Curves showing relationship between magnetic transition temperature (α-phase), resp. Curie temperature (γ-phase) and Ni content of an alloy.

For an alloy in the two-phase state the $\sigma - T$ curve shows two transition temperatures, the one by an inflexion in the curve and the second at the point where the magnetization finally falls to zero. Examples of such alloys are given in figure 3. Curve (a) shows a 16 % Ni alloy after 18 days at 525° C. From 140° C upwards this curve takes the form for a single-phase (α) alloy. The dotted section below this temperature is the result of extrapolation. The difference between this and the experimental result is the effect of a γ-component with a transition temperature at about 140° and is shown by the broken line curve (b). The Curie point of the γ-phase is obviously not well defined but from figure 2 we deduce that the Ni content of this phase is about $27 \pm 3 \%$ Ni. If the same alloy, however, be cooled to $- 180°$ C for even a short period before examination, the Ni-rich phase will undergo an irreversible change to the α-structure without changing its composition, provided it contains less than 33 % Ni. This change is an allotropic transformation of the lattice and as the temperature is much too low to allow diffusion to take place, it is justifiable to conclude that the composition of the phase cannot alter. The resulting $\sigma - T$ curve is that labelled (c) in figure 3. Here the transition temperature of the Ni-rich phase is sharp at 520° C so that its composition can be deduced more accurately from the graphs of figure 2, i.e. ~26 % Ni.

In both cases the iron-rich phase follows the same curve, thus indicating that neither its composition nor its amount is altered by the cooling in liquid air. This fact is an additional indirect proof that the composition or amount of the other phase present cannot be affected by its irreversible allotropic change of lattice structure. In figure 3 only a few of the numerous experimental points are included for the sake of clearness, but the others lie on the same smooth curve.
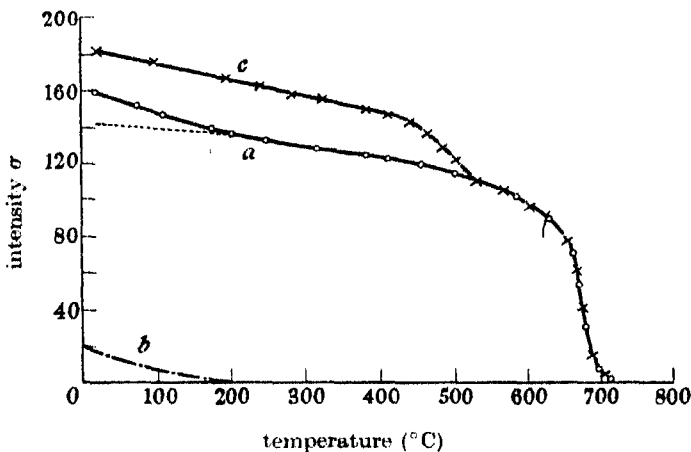
FIGURE 3. Intensity-temperature curves of 16% Ni alloys showing modifications produced by cooling a two-phase alloy to −180°C: (a) annealed 18 days at 525°C, (c) annealed and subsequently cooled to −180°C for 48 hr.
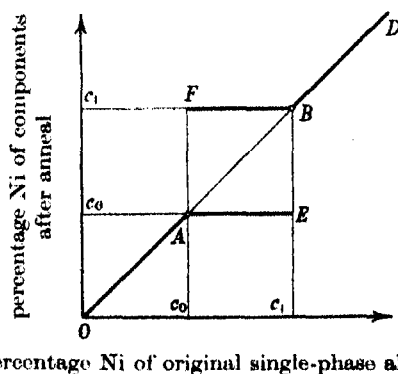
FIGURE 4. Theoretical form of relation between Ni content of alloys before and after anneal.

Let us now assume that we anneal a set of alloys at a temperature T until equilibrium is obtained. From the $\sigma - T$ curves of these specimens the composition of the components after anneal can be determined with the aid of figure 2 and a graph can be plotted giving this composition as a function of the composition of the original single-phase alloy. From the following considerations we should expect the form of this graph to be that shown in figure 4. In the region of com-

K. Hoselitz and W. Sucksmith

plete solubility of nickel in iron no segregation into two phases will have taken place and the alloy remains unchanged. The same will apply for the region of complete solubility of iron in nickel. In these cases the graph should therefore consist of straight lines through the origin, OA and BD respectively, making an angle of 45° with the axes. But in the two-phase region every alloy will have split up into a mixture of the two equilibrium phases. The nickel concentration of respective phases will be the same for all the two-phase alloys and only the amount of one phase relative to the other will be different in the various specimens. This will not affect the transition temperatures but only the relative intensities. The
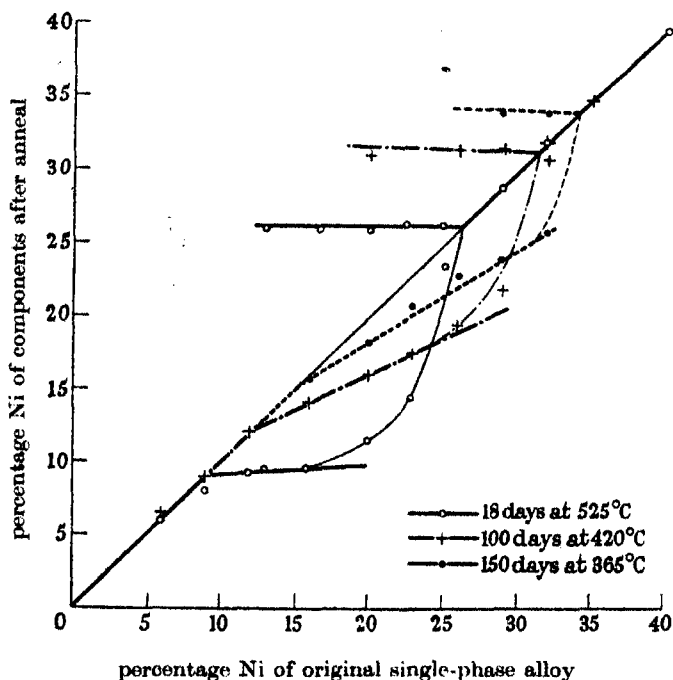
FIGURE 5. Experimental relation between Ni content of alloys before and after anneal.

graph in this region should therefore show two horizontal lines AE and FB, and the points where these intercept the line OD, A and B, give the composition c_0 and c_1 of the equilibrium phases at the temperature T. For the accurate experimental determination of c_1 it will be of advantage if the γ-phase is irreversibly changed into the α-structure, by cooling in liquid air, as demonstrated before.

This method of determining the equilibrium phase boundaries was employed for alloys annealed at 525° C for 18 days, at 420° C for 100 days and at 365° C for 150 days. In figure 5 these results are shown and thus the phase boundaries determined for 525° C at 9½ % Ni and 26½ % Ni, for 420° C at 12½ % Ni and 32 % Ni, and for 365° C at 15 % Ni and 34 % Ni respectively. The equilibrium diagram arrived at is shown in figure 6.

It will be seen that the phase boundaries differ somewhat from those given by Pickles & Sucksmith, shown dotted in figure 6. The main difference is in the α-boundary, but as mentioned before, the previous method did not lend itself readily to the accurate determination of the α-boundary at low temperatures. However as a check the α-boundary given by the above method was compared with that given by the earlier method of taking the room temperature intensities before and after annealing and observing the composition of the Ni poorest alloy which showed a fall of intensity after anneal. The α-boundary so determined is in good agreement with the newer method. The values found are between 7·6 and 9 % Ni for 525° C and between 12 and 13 % Ni for 420° C. These points are shown as crosses in figure 6. At 365° C the changes in intensity were too small to be observed at all.
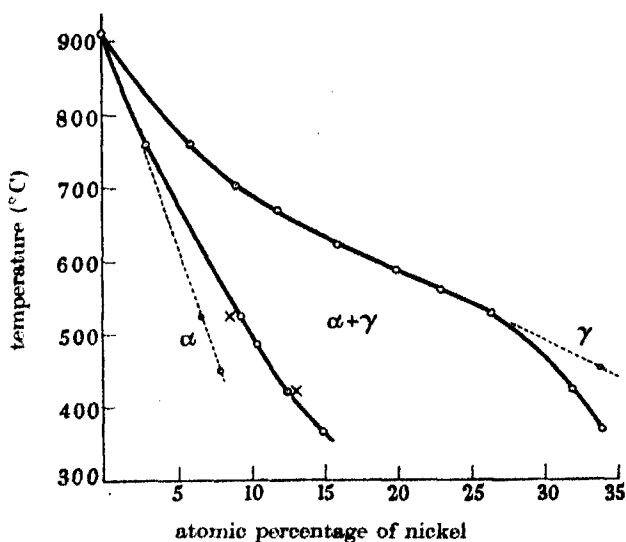
FIGURE 6. Equilibrium phase diagram of the iron-rich Fe-Ni alloys.

The γ-boundary, however, agrees with the present results, with the exception of the point for 450° C, which was thought to be at about 34 % Ni. Pickles & Sucksmith concluded that alloys annealed at 450° C were reversible between room temperature and liquid air temperature and that therefore the γ-phase contains more than 33 % Ni, the limit of irreversibility. In the light of recent evidence, however, it appears likely that the earlier preliminary heat treatments may not have been adequate in eliminating the effects of previous history. In the present work a very large number of alloys were brought into the quenched state with great care and were subsequently annealed at these low temperatures. All alloys annealed at 450° C, even at 420° C, were found to be irreversible and this is in good agreement with the fact that the γ-boundary at these temperatures is now located below the limit of irreversibility, i.e. below 33 % Ni.

MECHANISM OF SEGREGATION

In the curves of figure 5 a surprising feature is clearly borne out which, it is thought, throws light on the mechanism of phase segregation from the quenched state. It will be seen that although the lines in the two-phase field are in the main straight lines and allow a satisfactory determination of the equilibrium concentrations, they are not horizontal lines, as would have been expected. The alloys in this region have not split up into two phases with constant nickel percentages, but the concentration of the iron-rich phase is still a linear function of the concentration of the original single-phase alloy. The nickel-rich phase, however, is of constant percentage for all the two-phase alloys. This suggests that the process of segregation takes place by the γ-phase crystallizing out in its equilibrium percentage whilst the residue continuously changes its composition until equilibrium is finally reached.* In the usual case of a two-phase alloy both equilibrium phases crystallize out of the original lattice until the whole of this is used up. If this had also been so in the present case, one would have expected to find the presence of three constituents in specimens where equilibrium had not been reached. Although many specimens were far from equilibrium when examined, in most cases only two constituents were present. The few remaining cases were all in the neighbourhood of 25 % Ni, where there is reason to suspect a superlattice in the α-phase. Experiments were made to obtain definite information about a possible state of order in Fe_3Ni; the magnetic changes are so small that they fail to yield definite information. The fact that only two constituents were found in all other cases, supports the hypothesis put forward above with regard to the process of attainment of equilibrium.

In order to test this hypothesis still further, and also to investigate the rate of approach to equilibrium, a second series of experiments was carried out. Specimens with 16 % Ni were annealed for 2, 5, 9 and 18 days at 525° C. In each case only two constituents could be found. The nickel-rich phase had always the same composition, $26\frac{1}{2}$ % Ni, i.e. equilibrium percentage, and increased in amount, whilst the other constituent approached a percentage of $9\frac{1}{2}$ % Ni with increasing time of anneal. Also these experiments therefore confirm the suggested mechanism of phase segregation.

THE TIME-TEMPERATURE LAW

With the experimental material available it was thought possible to make an attempt at estimating the rate of attainment of equilibrium. Assuming that the speed of approach to equilibrium of the uniformly changing α-phase is proportional

---

* In a similar way, Gerlach (1936) showed the precipitation of one phase of Ni-Be in its final concentration with simultaneous uniform change of concentration of the residue. He also used the magnetic intensity at high fields, and later with Hammer (Gerlach & Hammer 1936) supplemented these measurements by experiments on the electrical resistance of the alloys. It appears likely that a similar process occurs in other alloy systems.

to the difference of its composition from equilibrium composition, similar to radioactive decay, we can write

(I) $\dfrac{dc}{dt} = K_{T_1}(c - c_0)$, where c is the concentration after the time t of anneal and

c_0 the equilibrium concentration at T_1. Assuming that the final state c_0 is constant, K_T is a function of the absolute temperature T only. We have thus, according to statistical considerations

(II) $K_T = A e^{-q/RT}$, where A is a constant of the dimension of a reciprocal time, the exact meaning of which is not yet correctly understood, and q is the activation energy in cal./mol., necessary for an atom to move in the lattice.

The assumption that the final state c_0 is constant is not exact, but in view of the slow change of the α-equilibrium percentage with temperature, we can assume c_0
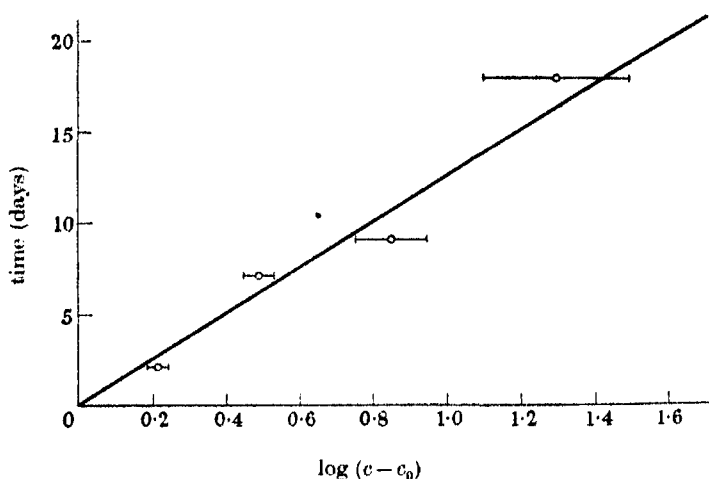
FIGURE 7. Approach to equilibrium of a 16 % Ni alloy annealed for various periods at 525° C.

to be constant at a first approximation. Furthermore, the crystallization of the γ-phase is probably dependent upon the presence of crystallization nuclei, which may have some influence on the rate of attainment of equilibrium, and this influence is not taken into account. The above formulae are therefore to be regarded as giving a crude description of actual conditions. They can be used to estimate the time necessary for an alloy in the quenched state to approach equilibrium to a desired degree.

From (I) by integration we see that when log $(c - c_0)$ is plotted against the time, the experimental points should lie on a straight line. Figure 7, showing the experimental results for the 16 % Ni alloy annealed for various periods at 525° C, is in reasonable agreement with this relation. This graph shows that K_T is really independent of the time of anneal and it remains now to test the relation of K_T as a function of T. From the experiments at the temperatures where the values of

$c - c_0$ could be determined with accuracy, K_T can be calculated using (I). The values of K_T thus obtained are:

T °C	K_T
525	0·22
490	0·12
420	$0·77 \times 10^{-1}$
365	$0·2 \times 10^{-1}$

These values allow the calculation of the 'half-value times', i.e. the times in which the composition of the α-phase of an alloy changes half way from original to equilibrium composition; thus according to (I) and (II), if $\log t$ is plotted against $1/T$ for any given $c - c_0$, in our case $\frac{1}{2}$, again a straight line would be expected. Figure 8 shows this graph for the four experimental temperatures. Here again the agreement of experiment and theory is satisfactory.
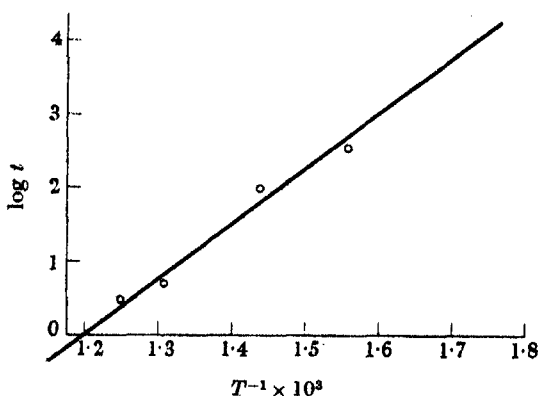
FIGURE 8. Half-value times of Fe-Ni alloys for various temperatures.

It is now possible to determine the values of A and q, which are $\sim 10^9$/day and 34,000 cal./mol. respectively. The value of A is smaller than that given by Dehlinger (1939) for Au-Ni, by a factor of 10^2, but in view of the sluggishness and slow rate of diffusion in the nickel-iron system, this is not surprising, since A is in some way connected with the frequency of an atom to change its place in the lattice, q is of the right magnitude. Jost (1937) gives values of 20,000 to 40,000 cal./mol. for the activation energy of intermetallic diffusion in mixed crystals with substitution lattices. Considering the approximate character of the suggested theory, the agreement can be regarded as satisfactory.

It might be of interest to remark that according to this treatment it is possible to throw light on two experimental questions. The first is the time of anneal necessary at 1000° C to bring an alloy into the single-phase state. It is found that from the quenched state the alloy would be in complete equilibrium after 4 hr. $[(c - c_0) \sim 10^{-20}]$, i.e. adequate time is allowed. It seems, therefore, that the traces of two phases which were found in the earlier work after heating times shorter than

24 hr. at 1000° C were due to the fact that the specimens in these cases were cooled too slowly. The second question of interest is that it is now possible to estimate the lower temperature limit at which any observable changes in the alloys can be expected within periods that can be used in laboratory experiments. For instance, at 300° C the half-value time would be about 40 years, at 325° about 10 years. Indeed, specimens annealed at 325° C showed no change after 6 months. It seems therefore reasonable to state that the lower practical limit of temperature, where the equilibrium diagram of the system can be investigated by ordinary annealing experiments, has been reached in the present work.

It should be emphasized once more that in all the experiments described above, the single-phase α-state was the original state of the alloys. Alloys that are left in the γ-state need much longer times to change towards equilibrium. Alloys that are annealed from an intermediate two-phase state show very complex characteristics, but in spite of this it is often possible to interpret their $\sigma - T$ curves in terms of the above mechanism of segregation.

REFERENCES

Bradley & Goldschmidt 1939 *J. Iron Steel Inst.* no. 2.
Dehlinger 1939 *Chemische Physik der Metalle und Legierungen.*
Gerlach 1936 *Z. Metallk.* 28, 80 and 183, 29, 124.
Gerlach & Hammer 1936 *Z. Metallk.* 29, 145.
Jost 1937 *Diffusion und chemische Reaktionen in festen Koerpern.*
Pickles & Sucksmith 1940 *Proc. Roy. Soc.* A, 175, 331.
Sucksmith 1939 *Proc. Roy. Soc.* A, 170, 551.

The crystal structure of 4 : 4'-dinitrodiphenyl [$C_{12}H_8(NO_2)_2$]

By J. N. van Niekerk, M.Sc., *University of Cape Town*

The crystal of 4:4'-dinitrodiphenyl [$C_{12}H_8(NO_2)_2$] is an example of one that can for purposes of measurement be referred to orthogonal axes, but which further investigation shows to be structurally monoclinic. The dimensions of the orthogonal unit cell, which contains two molecules, are $a = 3.77$A, $b = 9.56$A, $c = 15.39$A, and the space group is *Pc*.

The method of double Fourier series has been applied, and projections of the electron density on two of the axial planes have been made. The most important projection, that on the *bc* plane, has no centre of symmetry, and the phases of the Fourier components had to be estimated before summing the series. Although so far as the space group is concerned the molecules need have no centre of symmetry, it is almost certain that such centres exist.

Within the errors of experiment, the benzene rings are regular hexagons of side 1·41 A, the two rings in the molecule being co-planar with a common axis. The C-C link between rings is 1·42A. The C-N links do not lie in the plane of the rings, but make angles of about $\pm 22°$ with it. The two N-O distances in the nitro-group are not quite equal, being 1·14 and 1·21 A, the O-O distance being 2·00 A. The lines joining O to O in the nitro-groups are nearly parallel to the plane of the benzene rings, but lie about 0·42 A above and below it. The closest approach of O to CH in adjacent molecules is about 3·0 A, and that between CH and CH, 3·6 A.

A discussion of the packing of the molecules in the structure is given.

CRYSTALLOGRAPHY AND OPTICAL PROPERTIES

The material used in the investigations was prepared by Dr W. S. Rapson of the Chemistry Department of the University of Cape Town by the method of Gull & Turner (1929). Suitable crystals for examination by means of X-rays were obtained by slow evaporation of a solution in ethyl acetate. The crystals are light amber in colour, and are usually in the form of flat plates or thin elongated needles, some of which show the {011} prism faces round the zone [100]. Goniometer measurements gave the angles between such faces at 63° 30′ and 116° 30′. The directions of the crystal axes were chosen so that the angle between the (011) and (0$\bar{1}$1) faces was 63° 30′.

As an aid in guessing a provisional structure, the refractive indices of the crystal for light vibrating along the three principal axes were determined. Small crystals immersed in mixtures of α-monobromonaphthalin and paraffin, or for bigger refractive indices in methylene iodide and sulphur, were examined under a polarizing microscope. The concentration of the mixture was altered until the crystal became nearly invisible. The Becke line phenomenon was used as an indication of the direction in which the concentration had to be altered in any given case. By suitable manipulation of the crystals all three refractive indices were investigated in this way, and the values

$$\alpha = 1.49, \quad \beta > 1.78, \quad \gamma > 1.78$$

were obtained with sodium illumination for light vibrating parallel to the *a, b, c* axes respectively. The crystal is therefore optically negative, and very strongly doubly refracting.

THE UNIT CELL AND THE SPACE GROUP

A series of oscillation photographs was made with Cu $K\alpha$ radiation.* The dimensions of the unit cell obtained from these are

$$a = 3.77\,\text{A}, \quad b = 9.56\,\text{A}, \quad c = 15.39\,\text{A}.$$

The density of the crystal, as determined by the method of flotation, lies between 1·4 and 1·5 g./c.c. The unit cell must therefore contain two molecules of $C_{12}H_8(NO_2)_2$, corresponding to a true density of 1·45 g./c.c.

All the spectra recorded on the oscillation photographs fell accurately on both layer-lines and row-lines, and the films were indexed without ambiguity on rectangular reciprocal lattice nets, suggesting orthorhombic symmetry: however, for reasons given below it is clear that the crystalline symmetry is actually only monoclinic. General reflexions hkl of all types occur, showing that the unit cell having the above dimensions is primitive. Reflexions $h0l$ are present only if l is even. The structure therefore has glide planes of symmetry parallel to (010) with a glide $\frac{1}{2}b$. There are no other systematic absences.

That this structure cannot be orthorhombic may be shown as follows. If the crystal belongs to the orthorhombic class, there must be at least one, and possibly two, mirror planes of symmetry in addition to the single glide plane determined by X-ray measurements. Since the crystal contains only two molecules to the unit cell, both of which are accounted for by the glide plane, it follows that these additional mirror planes, if present, must be symmetry planes of the molecules themselves. Consider first the (100) planes. The observed spacing, 3·77 A, is about equal to the thickness of a benzene ring as determined from the structure of a number of aromatic compounds. The width of the rings is considerably greater than this (about 6·0 A), and so, if the (100) planes are symmetry planes of the molecule, they must coincide with the planes of the benzene rings. There are then two possibilities for the nitro-groups. Either the plane of these groups must lie in the (100) mirror planes, or the mirror planes must contain the nitrogen atom and be perpendicular to the line joining the oxygen atoms. The O-O distance in a nitro-group is 2·14 A according to the work of James, King & Horrocks (1935) on *para*-dinitrobenzene. Adding the distance 2·7 A for the radii of the oxygen atoms, we obtain a width of at least 4·8 A for the group, considerably greater than the (100) spacing of 3·77 A. The arrangement with the O-O distances perpendicular to the (100) planes is therefore ruled out, and the only possibility consistent with the molecular dimensions is that in which all the atoms of the molecule lie in the (100) planes. This means that all the atoms would scatter in phase for spectra of type $h00$. The 100 spectrum should be very strong, and the successive orders should diminish steadily in intensity. The 100 spectrum is actually weak in comparison with some of the stronger spectra recorded on the films, whereas on the assumption we are discussing it should have

* The observations for the $h0l$ spectra were made later using Fe $K\alpha$ radiation.

been the strongest o all. Moreover, 200 is absent and 300 very weak. We must conclude therefore that all the atoms cannot lie in the plane (100), and that this plane cannot be a mirror plane of the structure.

A similar argument, based on the molecular dimensions, and the size of the unit cell shows that the (001) planes cannot be mirror planes of symmetry of the molecule, and we are therefore left with the glide planes (010) as the only remaining symmetry planes of the structure. If this argument is correct, the structure is therefore not orthorhombic but monoclinic.

The results of this discussion, which is based principally on packing considerations, are supported by a closer examination of the crystals themselves, and of the rotation photographs. It is found that (1) the crystals do not show straight extinction when examined between crossed Nicols in a polarizing microscope, and (2) there is in some cases a marked difference between the intensities of the corresponding spectra hkl and $\bar{h}kl$, or of hkl and $hk\bar{l}$, while the intensities of hkl and $h\bar{k}l$ are always equal. At a later stage in the work, relative intensity measurements of all spectra of the form $h0l$ were made, and it was found that $h0l$ and $\bar{h}0l$ frequently differed greatly in intensity. These results are inconsistent with orthorhombic but consistent with monoclinic symmetry.

The elimination of the possibility of orthorhombic symmetry fixes the space-group as C_s^2 or Pc in the Hermann-Maugin notation, a group of the monoclinic hemihedral class. The rotation photographs show row-lines as well as layer-lines, and all the spots can be indexed accurately on rectangular reciprocal lattice nets. The crystal axes chosen are thus orthogonal, and although the crystal is structurally monoclinic, the monoclinic angle β, referred to them, is very nearly equal to $90°$.

PRELIMINARY ESTIMATE OF THE STRUCTURE

In attempting to estimate a probable structure for the crystal, it was assumed that both benzene rings lay in the same plane and had a common axis, so that the point midway between the two rings was a symmetry centre for the double-ring system. Provisionally, it was also assumed that the links from the benzene rings to the nitrogen atoms of the NO_2 groups were continuations of the axis of the ring system, and that this axis was also a two-fold symmetry axis of the nitro-groups; but rotation of the nitro-groups about the axis was assumed possible, in such a way as to maintain a symmetry centre at the point midway between the rings. Each ring was assumed to be a regular hexagon of side 1.40 A, in accordance with the results of Mrs Lonsdale (1929), Robertson (1936) and others. According to Dhar (1932) the C-C link in diphenyl has a length of 1.48 A, and this distance was assumed provisionally for dinitrodiphenyl. The size and shape of the nitro-groups assumed in this work were based on the results of James, King & Horrocks (1935) on *para*-dinitrobenzene.

With the assumed dimensions, solid cardboard models of the molecules were made to scale, and an attempt was made to pack them into the volume of the unit cell, with due regard to the symmetry conditions. In this preliminary work, the refractive

index measurements were of considerable help. The crystal has two very large refractive indices (> 1·78) for light vibrating parallel to the *b* and *c* axes, and one much smaller (1·49) for vibration parallel to the *a* axis. We may therefore infer that the directions *b* and *c* are nearly parallel to the plane of the rings, and that *a* is nearly perpendicular to the plane of the rings.

It was therefore assumed that the two benzene rings, and the two nitrogen atoms in the molecule lay parallel to the plane (100). In order to pack the molecules into the unit cell it was then necessary to incline the long axis of the molecule at an angle of about 35° to the *b* axis, and to tilt the nitro-groups so that the line joining the oxygen atoms made an angle of about 25° with the (100) planes.

The type of arrangement to which this reasoning leads may be seen from figure 2, which represents the projection of the final structure on the *bc* plane, viewed parallel to the *a* axis. More detailed work, based on intensity measurements, shows these preliminary ideas to be very near the truth.

STRUCTURE FACTOR CALCULATIONS

If the origin of co-ordinates is taken midway between the two glide planes (010) in the unit cell, the fractional co-ordinates of the two equivalent general points for the space-group *Pc* are

$$x, \ y, \ z; \ x, \ \tfrac{1}{2}-y, \ \tfrac{1}{2}+z.$$

It will be assumed in calculating the structure factors that the molecule has a centre of symmetry. The structure as a whole has no such centre; for this in conjunction with the glide plane would imply a screw axis, which does not exist. The centre of the molecule must therefore be displaced from the origin midway between the glide planes by a distance *p*, say, parallel to the *b* axis. There is no need to introduce any displacement parallel to *a* and *c*, for the distances between corresponding atoms in these directions are fixed by the translations of the cell and of the glide planes. The co-ordinate of an atom and of that symmetrically disposed to it relative to the centre of symmetry of the molecule may then be written

$$\pm (x, y - p, z),$$

and those of the corresponding atoms derived from these by the symmetry elements of the space-group

$$\pm (x, \tfrac{1}{2} - y + p, z + \tfrac{1}{2}).$$

The structure factor $F(hkl)$ for the general planes (hkl) may then be written

$$F(hkl) = A(hkl) + iB(hkl), \tag{1}$$

where, if $(k + l)$ is even,

$$\left. \begin{aligned} A(hkl) &= \ 4\Sigma f \cos 2\pi k p \cos 2\pi (hx + lz) \cos 2\pi k y, \\ B(hkl) &= -4\Sigma f \sin 2\pi k p \sin 2\pi (hx + lz) \sin 2\pi k y, \end{aligned} \right\} \tag{2}$$

and if $(k+l)$ is odd,

$$A(hkl) = -4\Sigma f \cos 2\pi kp \sin 2\pi(hx + lz) \sin 2\pi ky, \Big\}$$
$$B(hkl) = 4\Sigma f \sin 2\pi kp \cos 2\pi(hx + lz) \cos 2\pi ky. \Big\}$$ (3)

In (2) and (3) f is the appropriate atomic scattering factor of the atom whose co-ordinates are (x, y, z), and the summation is to be taken over all the atoms in one-half of the molecule.

THE FOURIER SYNTHESIS OF THE STRUCTURE

In determining the structure, the method of the double Fourier series, developed by W. L. Bragg (1929), has been used. If, as we have provisionally assumed, the molecule has a centre of symmetry, the projection of the structure parallel to the b axis on the plane (010) has also a centre of symmetry. The surface density of the projection $\rho(x, z)$ may therefore be written in the form

$$\rho(x, z) = \frac{1}{ac} \sum_{-\infty}^{+\infty} \sum_{-\infty}^{+\infty} F(h0l) \cos 2\pi(hx + lz),$$ (4)

$F(h0l)$ being given in magnitude and sign by $A(h0l)$ in equations (2) and (3).

The projections along the a and c axes have no symmetry centres, even in the case of the symmetrical molecule. The density $\rho(y, z)$ of the projection parallel to a on the bc plane is given by

$$\rho(y, z) = \frac{1}{bc} \sum_{-\infty}^{+\infty} \sum_{-\infty}^{+\infty} | F(0kl) | \cos \{2\pi(ky + lz) - \alpha(0kl)\},$$ (5)

where $\alpha(0kl) = \tan^{-1} \dfrac{B(0kl)}{A(0kl)}$, and $| F(0kl) | = \sqrt{(A^2 + B^2)}$. The projection $\rho(y, z)$ was determined first, since the provisional structure promised clear resolution of the molecules in the bc plane.

RELATIVE INTENSITY MEASUREMENTS

In order to evaluate the series for $\rho(y, z)$ it was necessary to determine the structure factors of the type $F(0kl)$ for all spectra of appreciable strength. For this purpose crystal prisms, elongated along the a axis, were used. One such prism, 0·3 mm. long and of nearly square cross-section with a side about 0·2 mm., was set for rotations about the a axis, the crystal being completely immersed in a beam of Cu $K\alpha$ radiation. For a crystal of this size, the absorption of the radiation could be neglected.

An ionization spectrometer and integrating photometer were not available, so that no absolute measurements of intensity could be made, and a set of relative measurements had to suffice. These were obtained by means of a photographically recording microphotometer, calibrated by a sector-wheel arrangement. Common spectra on successive films were used for standardizing one film with respect to the next, until the strength of each spectrum was given by at least six independent

determinations. The mean of these values was taken to represent the intensity of the spectrum.

The exact relationship between the relative strengths of the spectra measured in this way and their corresponding integrated intensities was not known; and this was a fundamental difficulty in the work, since it made a truly absolute comparison between observed and calculated values of $|F(hkl)|$ impossible. The assumption was made that the heights of the peaks on the photometer trace, properly calibrated, were proportional to the integrated reflexions of the corresponding spectra. This led to no inconsistencies, and was certainly a much better approximation to the truth than could be obtained by mere visual estimation. The way in which it proved to be possible to relate observed and calculated intensities when the nature of the structure was known showed that no great error had been introduced in this way.

The projection on the *bc* plane

Before evaluating the series, it was necessary to adjust the provisional structure until there was general agreement between the observed structure factors $|F(0kl)|$ and those calculated from the assumed atomic co-ordinates. As a first step, spectra of the type 00*l* were examined, for which the structure factor has the simple form

$$A(00l) = 4\Sigma f \cos 2\pi lz.$$

It was found possible to adjust the *z* co-ordinates so as to get a good general fit as regards rise and fall of intensities between observed and calculated values; and once this was done it was possible to get an idea of the absolute value corresponding to any observed strength of spectrum. The calculations were then extended to the 0*k*0 spectra, and an approximate value for the parameter *p*, which gives the displacement of the assumed centre of symmetry of the molecule from the origin in the direction of the *b* axis, was determined. Finally, spectra of the type 0*kl* were considered, and it was found possible to adjust the provisional co-ordinates in such a way as to obtain a fairly good general fit between the observed and calculated values of $|F(0kl)|$. With these adjusted provisional co-ordinates a complete set of values of $A(0kl)$ and $B(0kl)$ was calculated from equations (2) and (3), and from these the values of $\alpha(0kl)$ were evaluated for use in the Fourier series given by equation (5). In the calculations of $A(0kl)$ and $B(0kl)$ the values of *f* for C, N, and O tabulated by James & Brindley (1931) were used. No correction was made for thermal movements. Data for this purpose were entirely lacking, and in any case the motions involved are probably as much intramolecular as intermolecular and are consequently very difficult to allow for.

With the calculated values of $\alpha(0kl)$ and the observed values of $|F(0kl)|$, the series was evaluated, the method used being substantially that described by Lipson & Beevers (1936).

The contour diagram obtained from the final values of $\rho(y, z)$ for the projection parallel to *a* on the *bc* face is shown in figure 1. It may be interpreted by means of
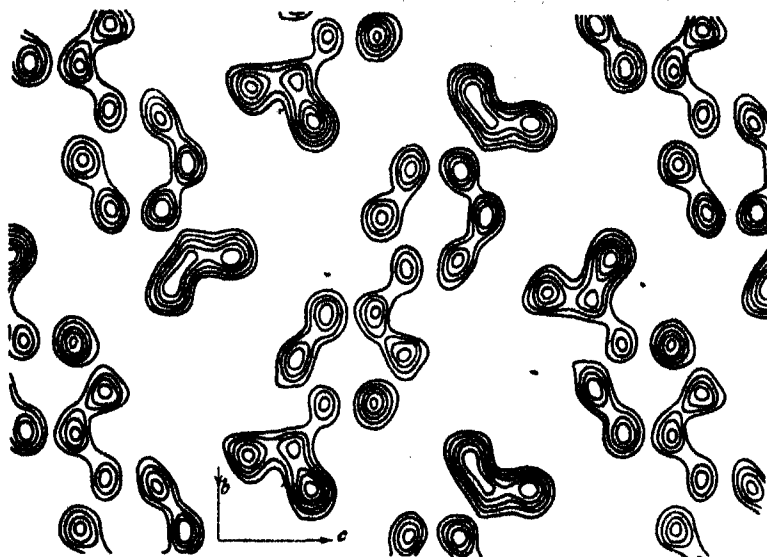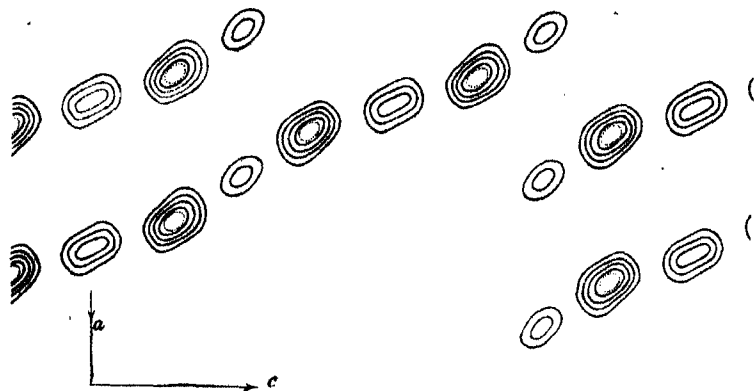
FIGURE 1. Fourier projection of the structure in direction a on the plane bc. Contours at intervals of 1·36 electrons per A². The outer line is the 2·72 electron contour.
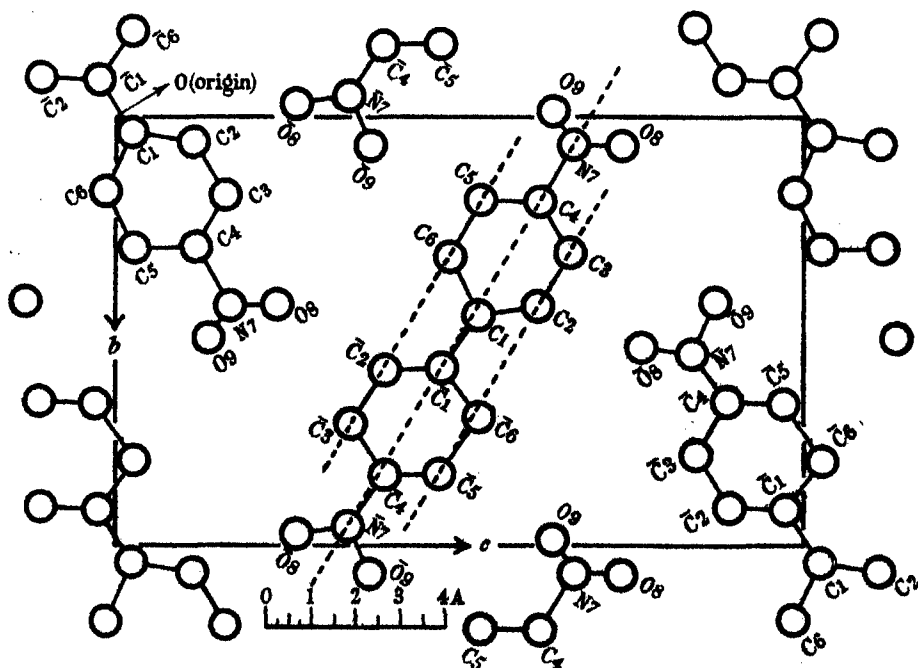
FIGURE 2. Projection of the structure in direction a on the bc plane.

the key diagram, figure 2. The projection gives a good outline of the molecules, and shows clearly the general scheme of packing. It suggests (a) that the two benzene rings in the molecule are co-planar, (b) that there is probably some tilt of the planes of the rings relative to the bc plane, but that this tilt cannot be very large, and (c) that the line joining the two oxygen atoms in the nitro-group is inclined to the bc plane, an assumption made provisionally in the tentative structure. It is not possible to determine with any accuracy from this projection the direction and amount of the tilts. The y and z parameters can be determined with considerable accuracy, but all the information that the projection gives about the x parameters is that these are relatively small, because of the approximately regular form of the projections of the benzene rings.

The strongest reflexions of the type 0kl are 008, 04 11, and 043. The projection accounts for this, since it shows that most of the atoms lie on or very near sets of planes with the corresponding spacings. With the co-ordinates obtained from the projection, the calculated intensities of 008, 04 11, and 043 are respectively 53; 57 and 51 % of the possible maxima. Assuming this to be correct, it is possible to get an approximate absolute standardization of the observed intensities, and this was of great value in determining the x parameters when the structure was projected parallel to b on the ac plane.

THE PROJECTION ON THE *ac* PLANE

Inspection of the projection on the bc plane shows at once that in neither of the other principal projections will there be any clear resolution of the atoms. It is evident also that of the two remaining projections, that parallel to b on the ac plane is likely to be the better. Moreover, with the assumption made as to the molecular symmetry, this projection has a symmetry centre, so that only the observed numerical values of the coefficients $|F(h0l)|$ need be used in evaluating the series. It is, however, necessary to get some idea of the probable value of the x parameters, in order that the correct sign may be given to the principal coefficients. The z co-ordinates of all the atoms were assumed to be known from the bc projection.

A number of preliminary tests showed that it was not possible to explain the observed h0l intensities by assuming the carbon atoms of the benzene rings and the nitrogen atoms of the nitro-groups all to lie in the (100) planes while an appropriate tilt was given to the line joining the oxygen atoms in the nitro-groups. It was soon evident that the plane of the benzene rings must be inclined to the (100) planes. There is a large difference in intensity between certain corresponding spectra of the type h0l and $\bar{h}0l$, particularly between 102 and $\bar{1}02$, 204 and $\bar{2}04$, 206 and $\bar{2}06$, which can be accounted for on no other assumption.

The clue to the ac projection is given by the very strong 102 reflexion. From the approximate absolute standardization of intensities described above, it appeared that the intensity of the 102 reflexion was 66 % of the possible maximum. This value

can be compared with those of the strong 008, 04.11 and 043 reflexions referred to at the end of the previous section. The atoms were in these cases found to be distributed roughly in sheets with the spacing about equal to that of the corresponding crystal planes. It could therefore also be inferred that the atoms of the molecules are arranged approximately in layers parallel to the (102) planes, with a spacing nearly equal to the (102) spacing. It was also found that 200 was absent, and that the intensity of 104 was 44 % of the possible maximum. These observations suggest a heavy concentration of scattering matter on or near the (102) and (104) planes to account for the strong 102 and 104 reflexions, and also a maximum distribution of atoms at intervals of roughly $\frac{1}{4}a$ to account for the 200 spectrum being absent.

Attempts were made to fulfil these conditions, while maintaining the assumed molecular symmetry, and to establish general agreement between observed and calculated values of $|F(h0l)|$. With the co-ordinates obtained in this manner, the signs of the principal coefficients of the series (4) for $\rho(x, z)$ were determined, and the series evaluated with the observed values of $|F(h0l)|$. The final contour diagram for the projection parallel to b on the face (010) is shown in figure 3, which can be interpreted by means of the key diagram, figure 4. The resolution of the atoms is very poor, and it is not possible to fix the x co-ordinates with any accuracy from it. Nevertheless, any x parameter which is finally chosen must be consistent with this projection.

It is evident from the bc projection that resolution in the ab projection will be worse than in the ac projection. No attempt was therefore made to determine the ab projection, and the final estimate of the parameters was made from the bc and ac projections.

<center>ESTIMATION OF THE PARAMETERS FROM THE CONTOUR DIAGRAMS</center>

In making the final estimate of the parameters the following assumptions were made: (a) that the benzene rings are co-planar, (b) that each is a regular hexagon of side $1 \cdot 41$ A, and (c) that the molecule as a whole has a centre of symmetry mid-way between the two benzene rings. All these assumptions are in themselves probable, and it has been found possible to give a consistent explanation of the contour diagrams and of the observed intensities without going beyond them.

From the bc projection it is evident that assumptions (a) and (b) are not far from the truth. It will be seen by following the dotted lines in figure 2 that the atoms marked C_1, $\bar{C}_4$, and $\bar{C}_6$ are displaced in the projection from their positions in a regular hexagon; but the displacements are not symmetrical, and have every appearance of being due to experimental error. There is no evidence of a regular distortion such as was suspected by James, King & Horrocks (1935) in *para*-dinitrobenzene. The shortening of the C–C link between the two benzene rings to $1 \cdot 42$ A as compared with Dhar's (1932) value of $1 \cdot 48$ A in diphenyl is probably due partly, if not wholly, to a displacement of the atom C_1. It must in fact be admitted that the accuracy of

the methods of intensity measurements used was not such as to permit the determination of the finer quantitative details of the structure. On the other hand, the general evidence from the projection supports the assumption of regular and co-planar rings. The y and z co-ordinates were estimated on this assumption.

FIGURE 3. Fourier projection of the structure in direction b on the plane ac. Contours at intervals of 3·45 electrons per A². The outer line is the 6·90 electron contour, and the dotted line the 18·98 electron contour. The molecule marked B2 (in figure 4) is not shown in this diagram.
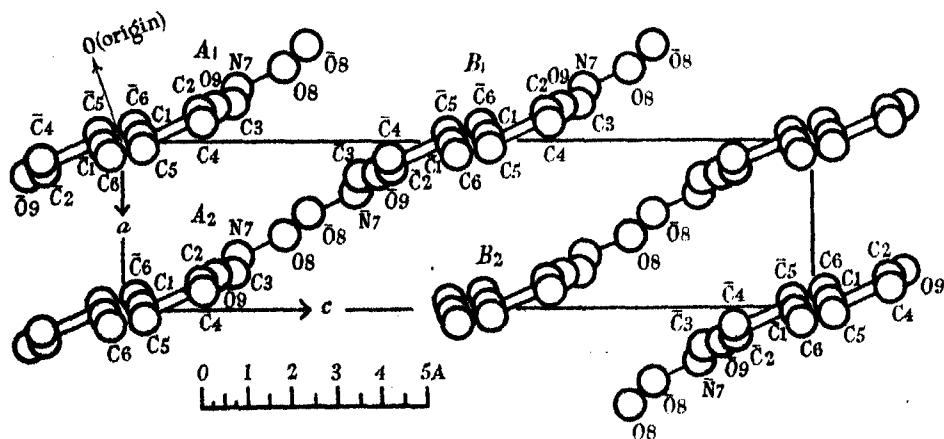
FIGURE 4. Projection of the structure in direction b on the plane ac.

In determining the x parameters and the tilt of the benzene rings relative to the (100) planes, assumptions (a), (b) and (c) were retained throughout. The atoms were given co-ordinates relative to axes lying in and perpendicular to the plane of the benzene rings, and various inclinations of these axes relative to the crystal axes were tried, until atomic co-ordinates, consistent with the ac projection and with the observed intensities, were obtained. The final agreement between observed and calculated intensities for spectra of the types $0kl$ and $h0l$ was reasonably good, and there were no bad inconsistencies. For reasons of economy of space no actual tables

showing this agreement are published. These have, however, been submitted with the paper, and can be seen by those interested.

DESCRIPTION OF THE MOLECULES AND OF THEIR ARRANGEMENT IN THE STRUCTURE

In describing the individual molecule, it is convenient to use a new set of axes, ξ, η, ζ, fixed relative to the molecule. The origin of this set of axes lies at the point mid-way between the two benzene rings, the ξ axis is perpendicular to the plane of the rings, and in the positive direction of a, the axis η along the long axis of the ring system and in the direction $C_1 C_4$, and the ζ axis is in the plane of the rings perpendicular to ξ and η. The co-ordinates, in A, of the atoms of the nitro-groups relative to these axes, as determined to give the best agreement with observations, are shown in table 1. The numbering of the atoms corresponds to that in the diagrams.

TABLE 1

Atom	ξ	η	ζ
N_7	-0.49	$+5.13$	$+0.25$
$\dot{N}_7$	$+0.49$	-4.97	-0.31
O_8	-0.48	$+5.69$	$+1.27$
$\dot{O}_8$	$+0.47$	-5.77	-1.28
O_9	-0.40	$+5.48$	-0.66
$\dot{O}_9$	$+0.44$	-5.61	$+0.77$

For the carbon atoms of the rings, the ξ co-ordinates are all zero, by assumption, and the η and ζ co-ordinates such as to form regular hexagonal rings of side 1.41 A. It will be seen that the co-ordinates of the two nitro-groups are very nearly symmetrical about the origin; indeed the departure from symmetry is within the limits of experimental error, and a symmetrical molecule may be assumed as consistent with observation.

The molecular dimensions, as based on the observations, are shown in table 2.

TABLE 2

C-C aromatic 1.41 A		N_7-O_8	1.21 A
C-C between rings 1.42 A		N_7-O_9	1.14 A
C-N 1.56 A		O_8-O_9	2.00 A

The figures for the nitro-group may be compared with those found by James *et al.* (1935) for *para*-dinitrobenzene, 1.25, 1.10, and 2.14 A respectively. The projections of the molecule on the planes $\xi\eta$ and $\eta\zeta$ are shown in figures 5a and 5b.

The arrangement of the molecules in the unit cell is best understood from figure 6, in which the structure is viewed parallel to the a axis, projected on the plane bc. The origin of the co-ordinates x, y, z lies at 0, and the traces of the glide planes are indicated by the discontinuous lines GP. We consider one of the two symmetrically
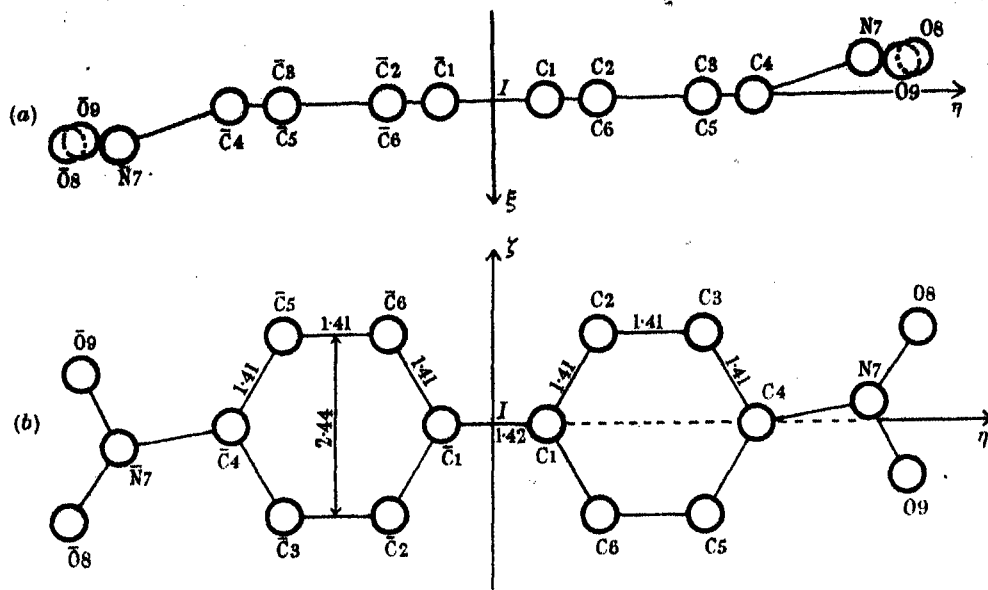
FIGURE 5. The molecule of 4:4'-dinitrodiphenyl.
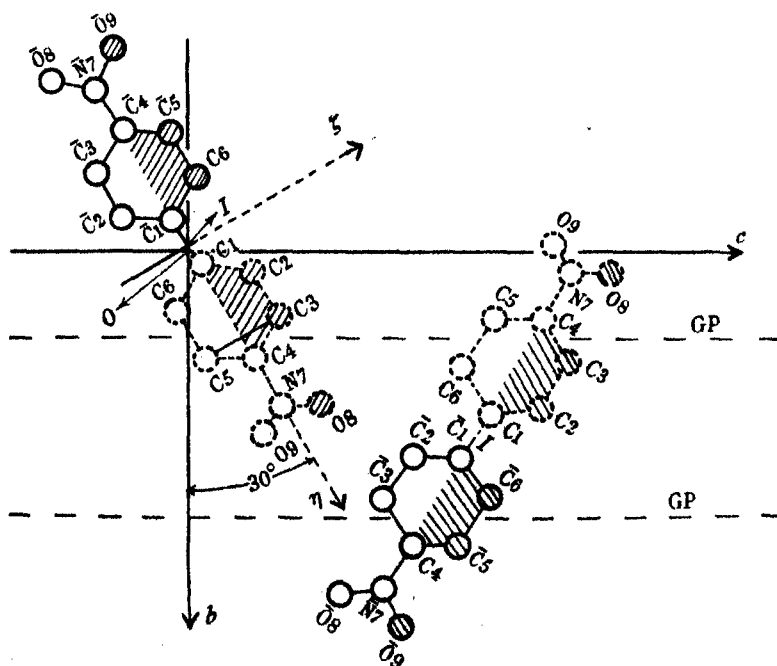
FIGURE 6. Projection of one molecule, and that derived from it by the operation of the glide planes *GP*, on the *bc* plane.

equivalent molecules. Its centre of symmetry I, the origin of the co-ordinates ξ, η, ζ fixed in the molecule, lies at the point $(0, -0·16, 0)$ relative to the axes a, b, c, and is thus displaced from 0 by a distance of $0·16$ A in the negative direction of b. The orientation of the molecule in the unit cell is perhaps most easily described in the following way. Assume as a first approximation that ξ is parallel to the a axis, while η and ζ lie in the bc plane, with the positive direction of η (IC_4 in figure 6) rotated away from the positive direction of the b axis through an angle of $30°$ towards the positive direction of the c axis. Next suppose the plane of the benzene rings to be rotated about IC_4 through an angle of $24°$ in such a direction as to depress the positive direction of ζ below the bc plane. The portions of the benzene rings so depressed are shaded in figure 6, while the unshaded portions lie above the bc plane after this tilt. Finally, suppose the molecule to be tilted through an angle of $7°$ about an axis through I, lying in the bc plane and perpendicular to the axis η, in such a direction as to displace C_4 below the bc plane in the negative direction of a. The direction of the projection of IC_4 on the bc plane is not altered by either of these tilts. The projections of lines such as C_2C_6 or C_3C_5 which are perpendicular to the η axis should make an angle of about $93°$ with the line IC_4 in the projection.

Those portions of the molecule depressed below the plane of the figure by the second tilt are shown dotted in figure 6. In this figure throughout, the dotted outlines refer to the depression produced by the tilt of the long axis relative to the plane bc, and the shading to that produced by the $24°$ tilt of the molecule about the long axis. The undotted and unshaded portions represent corresponding elevations. The corresponding elevations and depressions of the second molecule in the unit cell are at once obtained by considering the action of the glide planes, and are also indicated in this diagram.

In the bc projection, figure 2, the nitrogen atoms lie very nearly in the same line as the long axis $C_4\bar{C}_4$ of the rings. They therefore lie nearly in a plane through $C_4\bar{C}_4$ perpendicular to the bc plane. Calculations from the co-ordinates show that the link C_4N_7 makes an angle of about $22°$ with $C_4\bar{C}_4$. The lines joining the pairs of oxygen atoms in the nitro-groups are very nearly parallel to the planes of the benzene rings, but about $0·42$ A above and below them.

THE PACKING OF THE MOLECULES IN THE STRUCTURE

In discussing the packing of the molecules and their distances of closest approach, it is convenient to consider first of all the relationship to each other of the molecules whose centres of symmetry all lie in the same (100) plane. These molecules are of two types. In one type, that denoted by A, all have the same orientation and are derived from a single molecule by the b and c translations of the unit cell. Those of the other type, denoted by B, are derived from those of type A by the operation of the glide planes. Molecules of each type are shown in figure 7.

The closest distance of approach between neighbouring oxygen atoms is about $4·5$ A, much greater than the corresponding distance in inorganic structures, which

is about 2·7 A. The linkages holding the structure together are therefore probably not oxygen to oxygen. The closest links between adjacent molecules appear to be those between O and CH units. We may distinguish two types of O-CH linkages, (1) those between adjacent A molecules or adjacent B molecules, such as the links $O_9\bar{C}_6$, $O_9\bar{C}_5$, $\bar{O}_9 C_6$, $\bar{O}_9 C_5$, referred to as p linkages in figure 7, and (2) those between adjacent A and B molecules, such as the links $O_8\bar{C}_2$, $O_8\bar{C}_3$, $\bar{O}_8 C_2$, $\bar{O}_8 C_3$, referred to
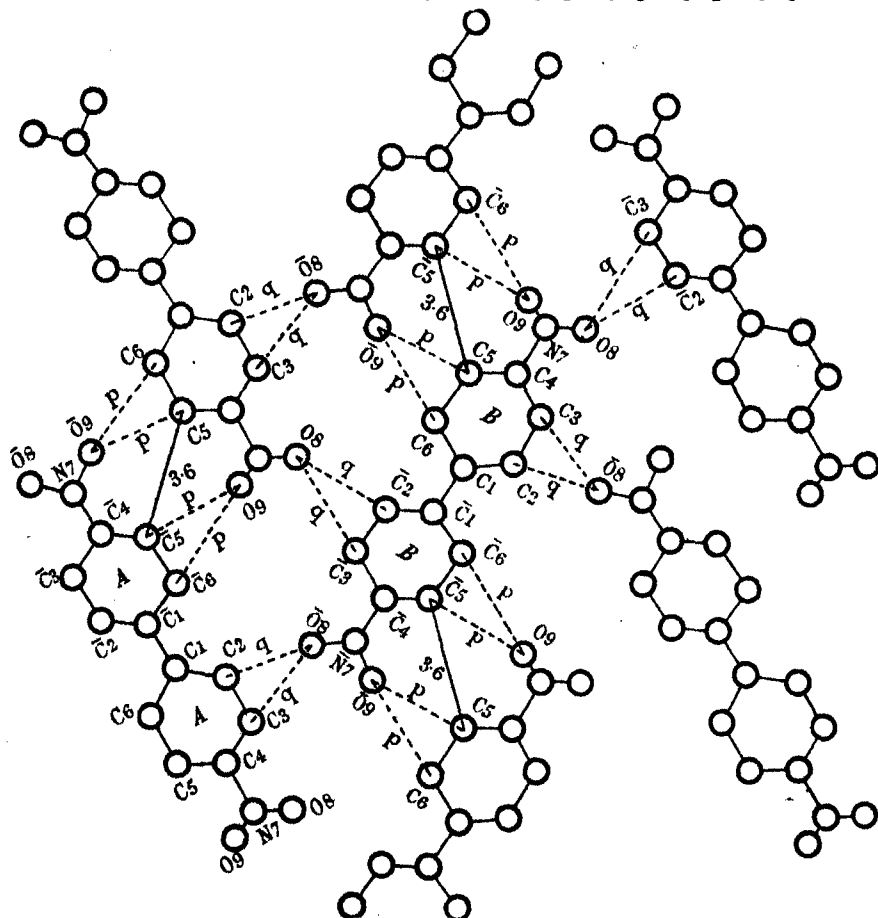
FIGURE 7. Projection of the structure on the plane bc viewed parallel to a, indicating the links p and q between O and CH units.

as q linkages in figure 7. The links of type p have values between 2·9 and 3·2 A, while those of type q have values between 3·4 and 3·6 A. Considering the molecular symmetry and the values of these links, it appears that all the p linkages tend to approach a mean value 3·0 A, and the q linkages a mean value 3·5 A. This suggests that the forces between two molecules of type A, or two molecules of type B, are greater than those between a molecule of type A and one of type B.

The closest approach of CH to CH in adjacent molecules is 3·6 A, and is indicated by the lines $C_5\bar{C}_5$ between two adjacent A or B molecules in figure 7. This value may

be compared with 3·60 to 3·72 A for naphthalene, 3·77 to 3·80 A for anthracene, obtained by Robertson (1933); 3·42 A for chrysene, obtained by Iball (1934); and 3·63 A for *para*-dinitrobenzene, obtained by James *et al.* (1935).

Reference to the molecule marked B in figure 7, shows that there are eight O-CH links of the type p with an approach of about 3·0 A to neighbouring B molecules in the same (100) plane. There are also eight O-CH links of the type q with molecules of type A, where the distance of approach is about 3·5 A. The values 3·0 and 3·5 A may be compared with the values 3·33 and 3·42 A found by James *et al.* (1935) for the O-CH links in *para*-dinitrobenzene.

The linkage between the molecules in adjacent (100) planes must now be considered. Figure 4 shows the structure projected on the (010) plane, and molecules of types A and B are again indicated. The a spacing is 3·77 A, so that corresponding atoms in molecules of type A or type B approach to within this distance of one another. The closest O-CH linkages between adjacent A molecules in different (100) planes is 3·9 A, and the closest link of any kind is that between O_8 and O_9 which is 3·4 A. The O-CH links between adjacent A or adjacent B molecules lying in different (100) planes are therefore longer than the corresponding links between molecules in the same (100) plane.

Reference to figure 4 shows however that the molecules form a series of sheets nearly parallel to the planes (102), suggesting that the principal linkages will probably lie between the molecules in these sheets. We find in fact that between the molecules A_2 and B_1 the links $O_8\bar{C}_2$, $O_8\bar{C}_3$, $\bar{O}_8 C_2$, $\bar{O}_8 C_3$, all approximate to the value 3·1 A, which may be compared with the links of type p between molecules in the same (100) plane. These shorter O-CH links between A and A or B and B molecules in the same (100) plane, and between A and B molecules in adjacent (100) planes, form a system of zigzag chains running through the structure, and probably represent the main binding forces in the crystal.

To sum up, every oxygen atom of the type O_8 approaches to within 3·5 A or less of six CH units, two such approaches being between molecules in the same (100) plane, and four between molecules in adjacent (100) planes. Each oxygen atom of the type O_9 approaches to within about 3·0 A of two CH units. Each NO_2 group, considered as a unit, thus makes eight such close contacts.

We have throughout this section used the term linkage in a purely geometrical sense, as denoting the distance of approach between atoms. It seems probable that the close O-CH links are in fact responsible for the greater part of the actual binding together of the crystal. The whole structure is very compact, the crystal possessing the fairly high density of 1·45 g./c.c., and is, for an organic crystal, comparatively hard. In this it resembles *para*-dinitrobenzene, and it is tempting to associate these properties with linkages of the type discussed between NO_2 groups and CH units.

due to Dr W. S. Rapson for preparing the crystals and for valuable information on the chemistry of this compound, and to Mr J. J. Duminy for help in connexion with intensity measurements and structure factor calculations. To the Research Grant Board of South Africa I am indebted for a grant.

REFERENCES

Bragg, W. L. 1929 *Proc. Roy. Soc.* A, **123**, 537.
Dhar, J. 1932 *Indian J. Phys.* **7**, 43.
Gull & Turner 1929 *J. Chem. Soc.* p. 494.
Iball, J. 1934 *Proc. Roy. Soc.* A, **146**, 140.
James, R. W., King, G. & Horrocks, H. 1935 *Proc. Roy. Soc.* A, **153**, 225.
James, R. W. & Brindley, G. W. 1931 *Phil. Mag.* (7), **12**, 81, 729.
Lipson, H. & Beevers, C. A. 1936 *Proc. Phys. Soc.* **48**, 772.
Lonsdale, K. 1929 *Proc. Roy. Soc.* A, **123**, 494.
Robertson, J. M. 1933 *Proc. Roy. Soc.* A, **140**, 79; **142**, 674.
Robertson, J. M. 1936 *Proc. Roy. Soc.* A, **154**, 187; **157**, 79.
Robertson, J. M. 1937 *Proc. Roy. Soc.* A, **162**, 568.

FERRIER LECTURE

Fatigue following highly skilled work

By F. C. Bartlett, F.R.S.

(Delivered 29 May 1941—Received 29 May 1941)

If the character and causation of fatigue following highly skilled work are to be understood, the first need is for the discovery of more relevant and experimentally controlled facts. Unfortunately, almost all the investigators who have attempted to study fatigue of this type have adopted methods taken over with very slight change from those which have proved valuable in the study of simple muscular fatigue. They have chosen elementary operations usually considered to require some 'mental' effort—such as easy calculations, word or colour recognition and naming and the like—have repeated these operations over and over again for long periods, and have tried to express the resulting fatigue in terms of the diminution in quantity or quality of the work done. The skill fatigue of daily life is not set up under such conditions. Routine repetition of simple actions is not a characteristic of any highly skilled work, and least of all of work having a strong 'mental' component. The operations involved here are marked by complex, co-ordinated and accurately *timed* activities. The stimuli in response to which these activities are set up are neither simple nor do they usually fall into an order of fixed succession. They have

the character of a field, or a pattern, which has become very highly organized, and may retain its identity in spite of a great diversity of internal arrangement.

It is possible to develop fully controlled experimental situations in which these realistic considerations have full play. When this is done the picture of fatigue following highly skilled work which emerges has certain strongly marked characters.

In such fatigue the 'standards' accepted and followed by the central nervous system unwittingly deteriorate. The operator tends to think that he is doing better work, because errors treated as significant all the time get wider and wider limits. Until a stage of great fatigue is reached, it is far more likely that the right actions will be performed at the wrong times than that the wrong actions will be performed. If accurate timing is insisted upon, gross mistakes of action may appear. The stimulus field splits up. Its pattern character alters. It becomes a collection of unconnected signals for action, with some of these predominant over all the others. Particularly, stimuli which are in the margin of the pattern, not closely organized with the central field, are ignored, 'forgotten', and serious lapses of specific reactions occur. There is a marked change in the effect of certain 'distracting', or additional stimuli. Sensations of bodily origin, in particular, become more pressing and insistent and affect the performance in ways peculiar to the tired operator. Side by side with all these changes go constant subjective symptoms. Verbal reports about any circumstances connected with known failure of performance become increasingly inaccurate, and errors are regularly projected upon objective conditions, or attributed to the interference of other people. There is a tremendous growth of irritability.

An attempt is made to discuss the light thrown by this picture upon the relation of high-level central nervous functions to simpler neuro-muscular mechanisms.

[*This lecture has been printed in full in Proceedings B,*
volume **131**, *pages* 247–257]

The structure of electro-deposited chromium

By William Hume-Rothery, F.R.S. and Malcolm R. J. Wyllie

The visual appearance and crystal structure orientation of electro-deposited chromium have been studied for deposits prepared at current densities from 50 to 3000 amp./sq.ft., and temperatures from 12 to 85° C, using a standard chromic acid bath containing 250 g. CrO$_3$ per litre, and a ratio of CrO$_3$ to sulphate ion of 100:1. Some measurements were also made at 95° C. A diagram has been constructed showing the effect of temperature and current density on the appearance and crystal orientation of the deposit.

The brightest deposits are characterized by a (111) preferred orientation, and no indication of any other fibre structure has been obtained. If the current densities are plotted against the logarithms of the temperature at which the brightest deposits are formed, a linear relation is obtained.

With increasing variation of the conditions of deposition (temperature or current density) from those characteristic of the brightest deposits, two effects are produced: (a) an increasing number of particles of purely random orientation are present in the deposit, and (b) the perfection of alignment of the particles of preferred orientation becomes less.

The residual stress present in electro-deposited chromium has been measured by the method of Stoney. If at a given current density the temperature of deposition is increased, the residual stress, which is contractile at first, rises sharply and may reach values as high as 110 tons/sq.in. This rise continues until the temperature is reached at which particles of preferred orientation first make their appearance. With further rise of temperature, the residual stress falls until it is practically zero at the temperature at which the brightest deposits are formed, and then rises again at higher temperatures.

The hardness of electro-deposited chromium has been measured for deposits prepared over the range 25–90° C at current densities of 500, 1000, and 1750 amp./sq.ft. The deposits of completely preferred orientation have the greatest hardness, and this maximum hardness is the same for all three current densities, although the temperatures of deposition at which the maxima occur are of course different. The temperature/hardness curves at different current densities can be almost exactly superposed by a mere shifting of the temperature scale, and it appears that the hardness is a property depending solely on the structure as revealed by X-rays, and is independent of the exact conditions under which the structure of a given type is produced.

1. Introduction

Although chromium can be deposited electrolytically from several different solutions, most if not all of the present-day chromium plating involves the use of chromic acid solutions containing small additions of sulphuric acid. It has been shown by Wright, Hirst & Riley (1935) that the body-centred cubic modification of chromium is always formed from solutions containing less than 18 % of chromium in the reduced (trivalent) state, irrespective of conditions of acidity, temperature, current density, etc., and this implies that the hexagonal modification is not to be expected from what may be called the standard chromic acid bath, containing 250 g. CrO$_3$ per litre, and a ratio of CrO$_3$ to sulphate ion of 100:1. When this electrolyte is employed, the nature, appearance, hardness, and other physical properties of the deposit are affected to a remarkable extent by the current density and temperature of deposition, but in spite of the wide use of chromium deposits, no systematic examination appears to have been made of the relations between the con-

ditions of deposition, and the structure and properties of electro-deposited chromium. Glocker & Kaupp (1924) examined one sample of a chromium deposit of 0·1 mm. thickness, and found it to possess a marked fibre axis with the [111] axis perpendicular to the plane of the deposit, but the conditions of deposition were not stated.* Wood (1931, 1935) investigated the structures of some deposits formed at low current densities, and interpreted abnormal relative intensities of the diffraction lines first (1931) as the result of particles of a definite shape relative to the crystal axes, and later (1935) as the result of preferred orientation, the nature of which was not determined. The only previous systematic investigation by X-ray methods appears to be that of Arkharow (1936), who showed that at the three current densities 20, 40, and 100 amp./sq.dm., lustrous deposits produced at 50–80° C possessed a [111] fibre structure, whilst grey matte deposits obtained at room temperature possessed a [100] fibre structure; the evidence for the latter was not very complete, and this fibre structure has not been confirmed by the present work. There appears to have been no previous systematic investigation of the relation between the structure, and properties of electro-deposited chromium, and even the visual appearance of the deposits has been examined systematically only at the lower current densities and temperatures. We have thought it of interest, therefore, to examine the visual appearance, and crystal structure orientation, over the whole range 50–3000 amp./sq.ft., and 12–85° C; some measurements were also made at 95° C. These measurements were supplemented by measurements of the hardness of deposits prepared over the temperature range 25–90° C, and at the current densities of 500, 1000, and 1750 amp./sq.ft., and of the residual stress in deposits prepared at 2000 amp./sq.ft. over the range 35–90° C. This extensive survey has shown how misleading conclusions were drawn from the observations at the lower current densities.

2. Experimental details

(a) *Electrolyte*. The electrolyte employed was of the chromic acid type, and contained 250 g. CrO_3 per litre, with the addition of sufficient sulphuric acid to give a CrO_3/SO_4 ratio of 100:1. The earlier experiments were made with 'Analar' chromic acid, and when this became unobtainable, use was made of chromic acid supplied by the Research Department, Woolwich. This material contained 0·06 % of trivalent chromium, and 0·21 % of sulphate, and was free from iron. Duplicate experiments indicated that identical results were obtained from the two varieties of chromic acid. The electrolyte in quantities of 2 l. was contained in a cylindrical glass beaker of diameter 17 cm. and height 10 cm., immersed in a thermostat controlled to within ± 0·2° C. Loss of water by evaporation was minimized by a clock glass cover. The electrolyte was usually changed when 0·005 % of the chromic acid had been reduced to metal, and in this way the accumulation of trivalent chromium (and of iron) in the electrolyte was kept very small.

* The deposits were prepared by the Grube process (U.S. Patent 1,496,845, 1924) using an aqueous solution of Cr_2O_3, $Cr(OH)_3$, and H_2SO_4, but the exact conditions are not stated.

(b) *Current supply.* The current for deposition was supplied by a 12 V accumulator of large capacity, and the circuit was provided with a mercury cup reversing switch which enabled the specimens to be anodically etched for a short time. The current was controlled by an adjustable rheostat, and was measured by means of a milli-ammeter or ammeter; the actual voltage across the terminals of the bath of course varied with the current density.

(c) *Preparation of specimens.* The preliminary experiments referred to on p. 335 were made with chromium deposited on steel wire of diameter 0·7 mm., but the remainder of the work has been carried out on specimens prepared from annealed steel sheet of 0·5 mm. thickness. An X-ray diffraction photograph of the steel showed that the crystal orientation of the material was entirely random. After cutting to the desired shape, the specimen was rubbed flat on no. 1 Hubert emery paper, and copper leads were soldered on at suitable places. These were protected from the solution by means of glass tubes, and cellulose varnish was used to stop off all parts on which deposition was not required. On placing in the solution, the specimen was made anodic, and a current of 200 amp./sq.ft. was passed for 75 sec. in order to remove any abnormal surface layer, and to degrease the specimen. The current was then reversed, and adjusted to the required value; this process usually required less than 3 sec.

(d) *Methods for the calculation of current density and for obtaining uniform deposits.* The mean current density was calculated from the mean current flowing, and the actual area plated. In the experiments with plated wires, a satisfactorily uniform deposit was obtained by the use of a lead anode bent into the form of a cylinder. With the plane specimens there was always a tendency for the current density to be higher at the edges of the specimen, and the deposits were not of uniform thickness. In order to allow a wide range of current density to be obtained, most of the work with flat specimens was carried out on rectangular sections of area 0·5 cm.². The centre of the deposit was used to obtain X-ray diffraction photographs, and the major axis of the elliptical X-ray beam at the surface of the specimen was found to be 2 mm. After the completion of the X-ray experiments, the specimen was mounted in Wood's metal, and was then sectioned, and two sections at right angles across the deposit were examined at a magnification of 750 in a Vickers Projection Microscope, and any specimens which were markedly irregular in thickness were rejected. In other cases, the thickness of the deposit was measured at intervals of 0·1 mm., and a diagram was constructed showing the thickness of the deposit as a function of the distance from the edge. If the irregularity of the deposit is slight, it is reasonable to assume that the efficiency of deposition is constant over the small range of current densities concerned, and in this way by measuring the area under the curve, it was possible to estimate the true current density for the portion submitted to the X-ray beam.

The results obtained by the above methods were quite consistent, but as an additional precaution the results for some of the more critical deposits were checked against further deposits prepared on circular disks of diameter 2·5 or 1·7 cm.,

according to the current densities used. These disks were placed inside cylindrical bakelite troughs, arranged as shown in figure 1. The bakelite is unattacked by the electrolyte, and the device enabled very uniform deposits to be obtained. This apparatus was only perfected towards the end of the present work, after a number of other devices of the 'robber bar' type had been tried. It was found that the more critical results obtained in the earlier work, and corrected as described above, were exactly confirmed by the more uniform deposits obtained with the apparatus of figure 1, and it was, therefore, not thought necessary to repeat the whole of the earlier results.
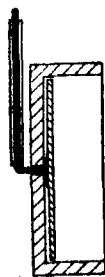
FIGURE 1
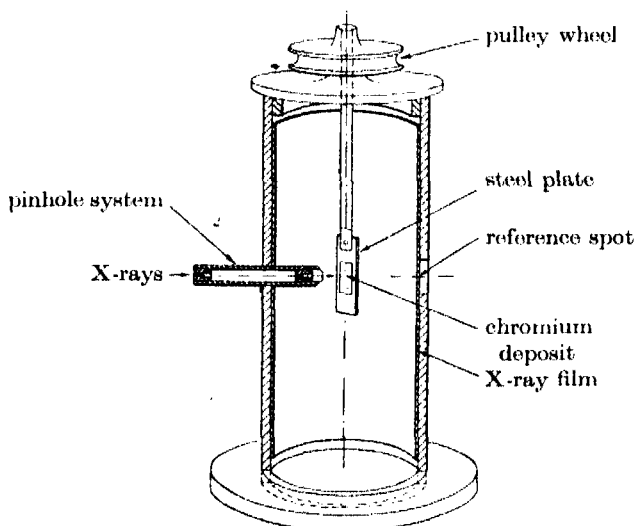
FIGURE 2. Section through pinhole system

(e) *X-ray technique.* The preliminary work was done with wire specimens in a Debye-Scherrer camera, whilst the detailed work was carried out with plane specimens which were mounted in a specially designed X-ray camera of the type shown in figure 2 which is self-explanatory. The angle between the plane of the specimen and the incident X-ray beam can be varied by rotating the specimen holder, and the diffraction lines and spots are recorded on a film of size 12×9 cm. which is bent into a cylinder inside the camera. The construction of this camera, the methods of measurement, and the interpretation of results are being described by one of us (M. R. J. W.) in a separate paper. Figure 3 A shows a typical film of a deposit with completely random orientation, and figures 3 B, C and D are from three deposits with different degrees of preferred orientation. When compared with an X-ray film from electro-deposited nickel, the films from the chromium deposits were found to be characterized by very intense general scattering, and many photographs were therefore taken through a screen of aluminium foil which greatly reduced the general scatter.
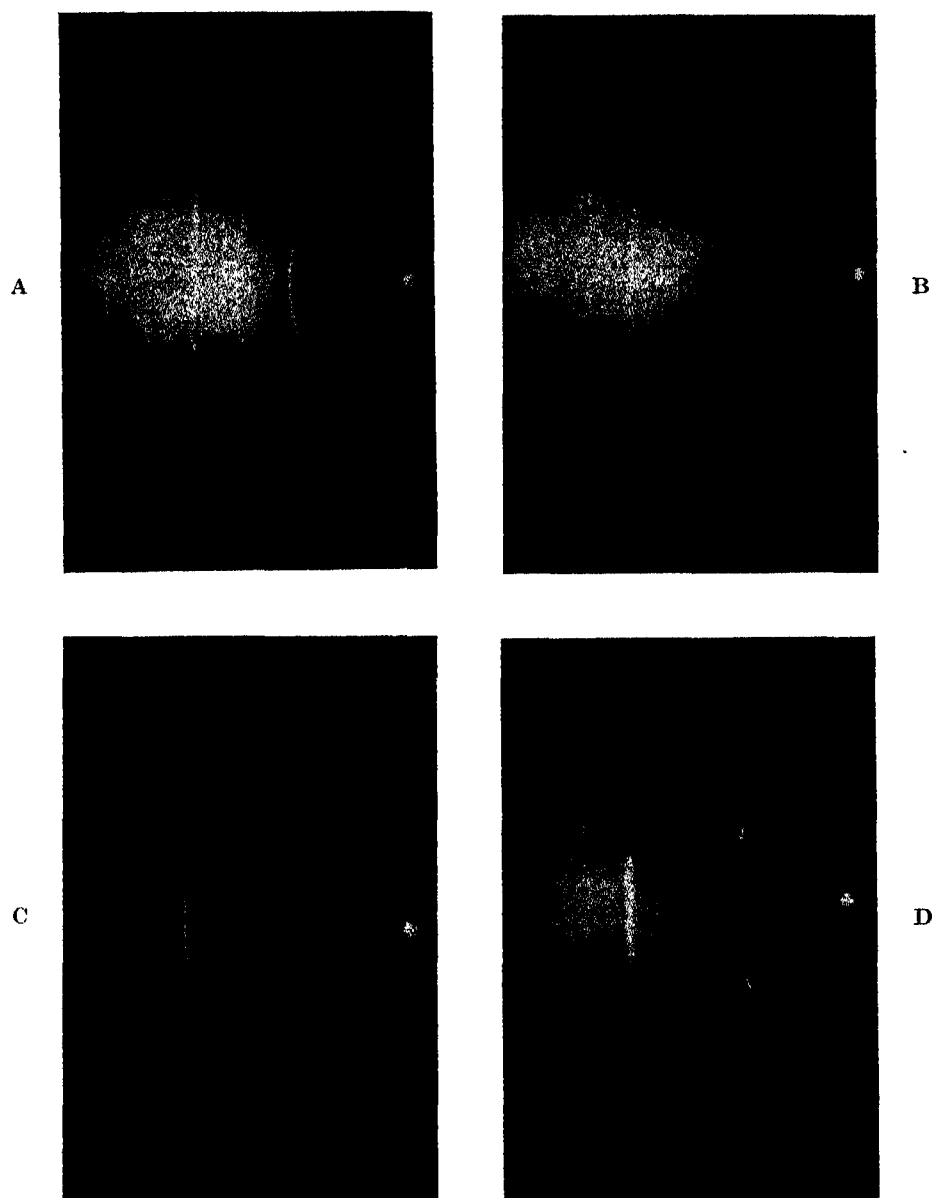
3. EXPERIMENTAL RESULTS

(1) Effect of temperature and current density on appearance of deposits

Our first work was with deposits made on steel wires, and we have to thank Mr E. E. Jarvis for making an extensive survey of the subject which has been of

great use in leading to the final results. It was found that the curvature of the wire led to difficulties in interpreting the X-ray diffraction patterns, and the deposits were therefore repeated on plane specimens, and the results are summarized in figure 4 in which the co-ordinates are temperature and current density, and the nature of the deposit is indicated. The field of this diagram may conveniently be divided into four zones with the following characteristics:
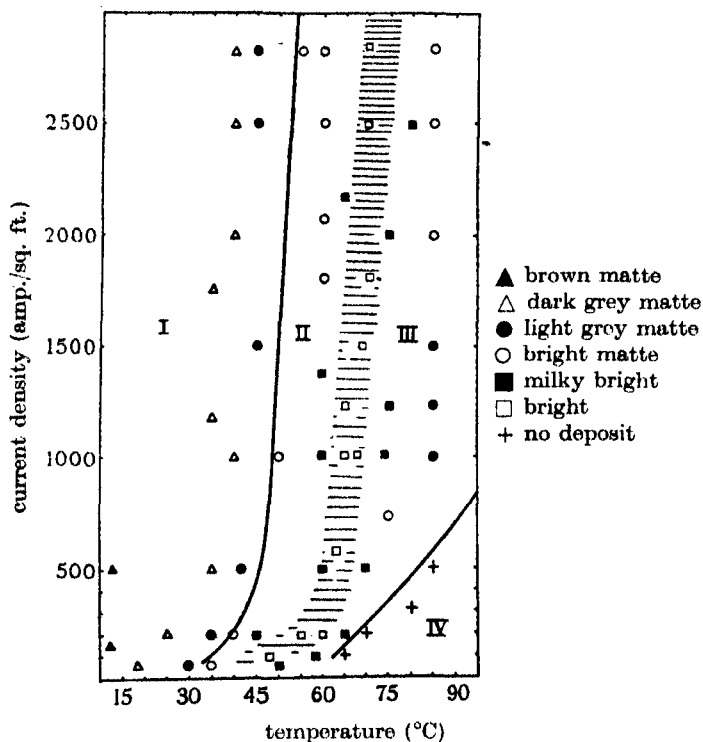
FIGURE 4

Zone I. The deposits in this zone are not lustrous, but are dark matte in appearance, and at the lower temperatures are of a dark brown or chocolate colour. At the lowest temperatures, they become more and more flaky as the current density increases. This last observation confirms the work of Shishkin & Gernet (1927).

Zone II. In this zone the deposits are definitely lustrous, and they become more and more bright on passing from the left to the right of zone II in figure 4. The brightest deposits occur at the boundary between zone II and zone III, and show specular reflexion. Although not absolutely sharp, the boundary between zones I and II is much more definite than that between zones II and III, and it will be seen from figure 4 that with current densities greater than 800 amp./sq.ft. the boundary between zones I and II only varies by a few degrees, although below 500 amp./sq.ft. the temperature of the boundary varies rapidly with current density.

Zone III. In this zone the deposits become less bright and develop more of a milky or matte appearance. The boundary between zones II and III is not a very definite one, and is therefore represented by a shaded area instead of by a single line, and this resembles the boundary between zones I and II in that, at the higher current densities, the effect of temperature is predominant. The upper part of the boundary between zones II and III in figure 4 is not, however, exactly vertical, and it is perhaps significant that if the current density is plotted against the logarithm of the temperature at which the brightest deposits are formed, a linear relation is obtained, suggesting that there is an exponential relation between the two variables.

Zone IV. In this zone, no deposition occurs, in agreement with the views of Blum (1932) and Biriükoff (1936) that a critical current density is required before deposition occurs.

Diagrams resembling figure 4 have been shown over a limited range by earlier workers (see for example, Blum 1932), but the present work appears to be the first to have covered such a wide range of temperature and current density, and is the first to show the predominant effect of temperature at the higher current densities.

(2) *The crystalline nature of the deposits*

As will be appreciated from what follows below, some of the deposits were characterized by a preferred orientation or texture, and the degree of texture was sometimes profoundly influenced by the thickness of the deposit. For this reason we have made first a survey of the crystalline nature of deposits of a standard thickness of approximately 10μ, and the effects of increasing thickness are dealt with later.

With deposits of thickness approximately 10μ, the deposits formed in zone I, whether of the brown or grey matte types, always possess a body-centred cubic structure, and no trace of the hexagonal modification has been found. The X-ray diffraction lines are reasonably sharp and well defined, and the Debye-Scherrer rings are continuous without any sign of preferred orientation. These results indicate that the deposits in zone I have a crystal size of the order 10^{-3}–10^{-5} cm., and that the orientation is perfectly random. No evidence was obtained for the existence of the (100) preferred orientation claimed by Arkharow (1936).

The deposits in zone II of figure 4 are all characterized by a preferred orientation in which the (111) planes of the crystal tend to be parallel to the plane of the deposit. At the left-hand side of zone II in figure 4 there is still a number of particles with random orientation, and the X-ray films show continuous Debye-Scherrer lines, of non-uniform intensity (cf. figures 3 B, C), the darkened portions* being in the positions required by a (111) orientation, but with a certain amount of variation in the exact alignment. As the right-hand side of zone II is reached, the continuous Debye-Scherrer lines vanish, showing that the number of particles with random orientation has become negligible, and at the same time the darkened arcs become

---

* The dark and light portions are reversed in reproduction.

shorter (cf. figures 3 C, D), showing that the orientation is much more perfect. At the boundary between zones II and III where the deposits are brightest, the departure from an exact (1 1 1) orientation is of the order $\pm 7\frac{1}{2}°$, whilst at the left of zone II the variation is of the order $\pm 20°$. Table 1 summarizes the data obtained in this connexion. Movement from the left to the right of zone II thus produces two distinct effects, namely, (1) a decrease in the relative number of crystals with a purely random orientation, and (2) an increase in the perfection of the exact alignment of the orientated crystals. Some of the bright deposits on wires gave very diffuse lines in the Debye-Scherrer camera, and application of the usual formulae connecting the particle size with the half-breadth of the lines suggested a particle size of the order 10^{-7} cm. We think, however, that this figure must be accepted with the greatest reserve, because the whole problem of the line widths of the diffraction lines from these deposits is extremely complicated. Apart from the question of the validity of the particular equation used for the calculation of particle size, there are also the possibilities of (1) abnormal lattice spacings due to co-deposited hydrogen, (2) effects of high stress in some deposits (see p. 339), and (3) the effect of superposition of the continuous Debye-Scherrer line (due to the random particles) on the arcs due to the particles with preferred orientation.*

TABLE 1

current density amp./sq.ft.	temp. °C	appearance	divergence from exact (111) orientation	relative proportion of random crystals
2500	45	dark grey matte	...	all
	50	bright matte	----	bulk
	60	bright matte	$\pm 16\frac{1}{2}°$	considerable
	65	milky bright	$\pm 8\frac{1}{2}°$	trace
	70	bright	$\pm 7°$	none
	75	milky bright	$\pm 9\frac{1}{2}°$	trace
	85	bright matte	$\pm 17°$	considerable
1000	35	dark grey matte	...	----
	50	bright matte	...	very considerable
	55	bright matte	$\pm 14\frac{1}{2}°$	considerable
	65	bright	$\pm 6°$	none
	75	milky bright	$\pm 15°$	considerable
	85	light grey matte	----	very considerable
200	35	light grey matte	...	---
	45	bright matte	$\pm 15°$	considerable
	55	bright	$\pm 9\frac{1}{2}°$	none
	60	bright	$\pm 9°$	none
	65	milky bright	$\pm 12°$	considerable

On passing from zone II into and across zone III the relative number of crystals with purely random orientation increases continuously, and at the same time the

* Calculation suggests that if the deposits examined by Wood (1931) possessed the same characteristics as our own, the grain size given by the latter refers to the particles of random orientation, and not those of preferred orientation.

perfection of the exact alignment of the crystals with preferred orientation decreases. At 85° C the deposits still show a certain degree of preferred orientation, but at 95° C they are completely random. There is also a remarkable increase in the sharpness of the diffraction lines on passing from 70 to 95° C, which suggests that the grain size of the deposited metal is increasing.

The effect of thickness on the degree of preferred orientation depends markedly on the nature of the deposit. In the case of the brightest deposits at the boundary between zones II and III, a highly preferred orientation is rapidly acquired, and in deposits of thickness 1μ the preferred orientation is fully as great as in a film of thickness 10μ. In those deposits which consist of a mixture of particles with random and preferred orientations, the effect of the thickness of the deposit is very marked, and as the deposit becomes thicker the degree of exact alignment of the particles with preferred orientation becomes more perfect, although they are still accompanied by roughly the same proportion of random particles. Thus with a deposit of thickness 36μ produced at 85° C and 1000 amp./sq.ft., the departure from an exact (111) orientation was only about one-third as great as for a deposit of thickness 10μ produced under the same conditions.

(3) *The existence of stress in the deposits*

It is well known that many electro-deposited metals are deposited in a stressed condition, but no previous systematic investigation appears to have been made of the magnitude of the residual stress in chromium deposits, or of its dependence on the conditions of deposition, although for nickel Stoney (1909) has shown that stresses as great as 19·2 tons/sq.in. were reached. This work was carried out by depositing nickel on one side of a flexible steel strip, and measuring the direction and magnitude of the curvature of the strip after the deposition was complete. We have used the same method for the study of residual stress in chromium, deposited on one side of thin steel strip supplied by the Research Department, Woolwich. This material was 5 mm. wide, and 0·35 mm. thick, and possessed very uniform properties. It was cut into lengths of 13 cm., and one side was completely stopped off with cellulose varnish, and the other side was stopped off for a distance of 3 cm. from one end, this end being held securely in a suitable holder. A length of 10 cm. was thus left for deposition, and the movement of the free end relative to a fixed glass pointer was measured. From this deflexion, the stress which was always contractile was calculated by the method described by Stoney (1909).

Figure 5 shows the results obtained for deposits prepared at 2000 amp./sq.ft., and at temperatures from 35 to 95° C. From this it will be seen that, on increasing the temperature, the stress increases rapidly to a maximum value of 110 tons/sq.in. at the boundary between zones I and II of figure 4. On entering zone II the stress falls rapidly as the degree of preferred orientation increases, and becomes almost zero at the boundary between zones II and III where the brightness and degree of preferred orientation are greatest. On entering zone III the stress rises to a much less

pronounced maximum at about 85° C. This work therefore establishes conclusively that the bright deposits with preferred orientation have the lowest residual stress as measured by Stoney's method.

The results of figure 5 refer to the residual stress in the deposit as a whole, but information regarding the distribution of stress through the deposit can be obtained by dissolving away the steel basis with nitric acid, and examining the direction of curvature of the chromium film. Working in this way we have found that the dark
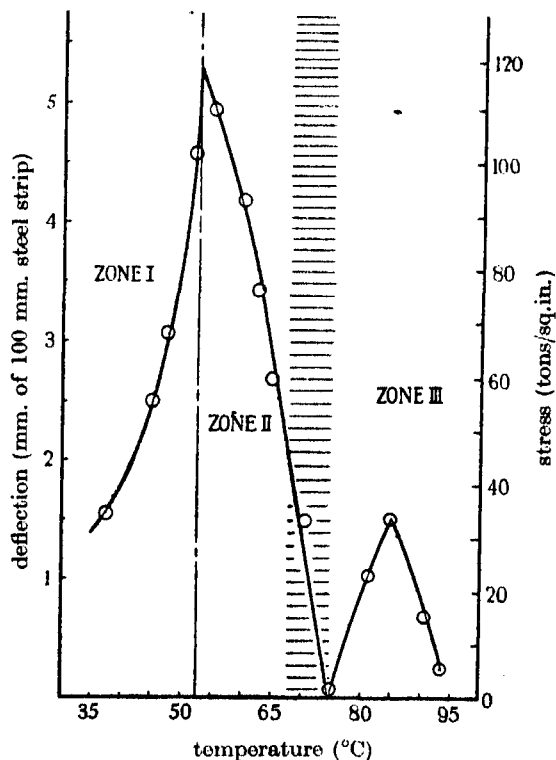
FIGURE 5

matte deposits formed at the low-temperature side of zone I show little or no tendency to curl. The bright matte deposits formed at the left of zone II show a clear tendency to bend so that the convex side is that originally facing the steel. This shows that the tension is greater in the outer layers of these deposits, which consist of a mixture of particles with preferred and random orientations. In contrast with this, the bright deposits tend to bend slightly so that the concave side is that originally facing the steel. This shows that for these deposits which consist entirely of particles with preferred orientation, the tension is somewhat greater on the inner layers of the deposit.

The hardness of electro-deposited chromium

The hardness of electro-deposited chromium is known to be very great, but the only extensive previous investigation of the relation between the hardness and the conditions of deposition appears to be that of Cymboliste (1938). This work was confined to the range 35–75° C, and 150–800 amp./sq.ft., and showed clearly that at a given current density of deposition, the hardness at first increased with increasing temperature of deposition, rose to a maximum, and then decreased at the higher temperatures. The absolute numerical values of Cymboliste are open to doubt, because although the Vickers Pyramid Indenter was used,for the tests, the results are only given in Brinell units, obtained by means of an unstated conversion table, and it is known that such conversion methods are unreliable for very hard substances. Although it is generally known that the bright deposits of chromium are the hardest, there appears to have been no previous systematic investigation of the relation between the hardness of the deposits and their structure, and we have therefore made a comprehensive investigation for deposits prepared over the range 25–90° C, at current densities of 500, 1000, and 1750 amp./sq.ft. respectively.

For this purpose deposits were prepared on standard $\frac{1}{2}$ in. diameter Hoffmann steel balls,* electrical connexion being made by means of a glass-encased copper wire. By using cylindrical lead anodes, a perfectly uniform deposit was obtained; this method was suggested to us by Mr A. W. Hothersall, and has proved most satisfactory. The hardness tests were made with a standard square-based diamond pyramid indenter, using a load of 6·1 kg. applied by means of a simple lever machine, the use of which has already been described by Frye & Hume-Rothery (1942).

Preliminary tests were made with deposits prepared on the hard steel balls, and also on balls softened by annealing. These experiments showed that the measured hardness was independent of the hardness of the basis metal, provided that the thickness of the deposit was from 10 to 13 times the depth of the indentation. All measurements were therefore carried out on deposits at least 15 times as thick as the depth of the indentations,† and all the usual precautions for accurate hardness testing were adopted, including that of allowing for the fact that the impression was on a curved and not on a plane surface.

The results obtained are shown in figure 6 and are of considerable interest. At each current density, increasing temperature produces first a marked increase in the hardness, which reaches a well-defined maximum at the boundary between zones II and III of figure 4, and then falls rapidly as zone III is entered. The deposits of completely preferred orientation have thus the greatest hardness, and this maximum hardness is almost the same for all three current densities, although the temperatures of deposition at which the maxima occur are of course different. Further, the complete

---

* These are made to rigid standards for ball bearings, and proved very satisfactory and uniform.

† Although a full investigation was not made, it seems certain that the hardness values are not affected by variations in thickness between 15 and 30 times that of the indentation.

curves at all three current densities can be almost exactly superposed by a mere shifting of the temperature scale, and it appears that *the hardness is a property depending solely on the structure of the deposit as revealed by X-rays, and is independent of the exact conditions under which a structure of a given type is produced.* Thus at the boundary between zones I and II in figure 4, where the preferred orientation begins, the hardness is 830–850 v.p.h. for all three current densities, although the temperatures of deposition at which this type of structure is produced are quite different. Again, the grey matte deposits produced in zone III appear to be characterized in every case by a hardness of about 500 v.p.h.
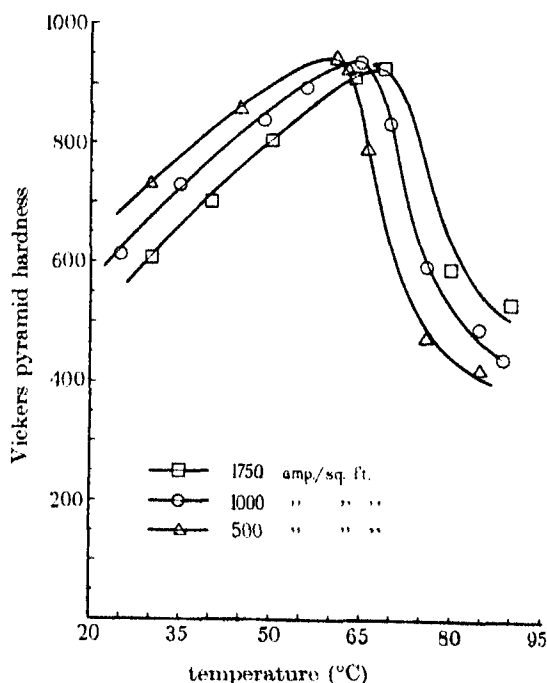
FIGURE 6

4. CONCLUSION

The preceding sections have dealt with the appearance, structure, residual stress, and hardness of electro-deposits of chromium prepared under very strictly controlled conditions. In the course of this work, some other points have been investigated, and may be referred to briefly, since they refer to deposits prepared under identical conditions.

(1) *The effect of the basis metal.* The present work does not support the view that the orientation of the deposit, at the thicknesses concerned, is controlled by the basis metal, since it shows that deposits of highly preferred orientation are obtained on steel strip of random orientation. As a further test of this point, a thin (1μ) deposit of chromium was made on an etched nickel base, but although the nickel possessed

a marked (100) orientation, the chromium had the unusual orientation. These facts suggest that the orientation of electro-deposited chromium, at the thicknesses concerned, is a property of chromium itself, and not that of the base.

(2) *Efficiency data.* Figure 7 A shows efficiency data we have obtained at 2500, 1750, and 1000 amp./sq.ft., and figure 7 B shows similar curves obtained by Schneidewind, Urban & Adams at current densities from 50 to 800 amp./sq.ft. Although the solution used by Schneidewind, Urban &'Adams (1928) differs from our own, it
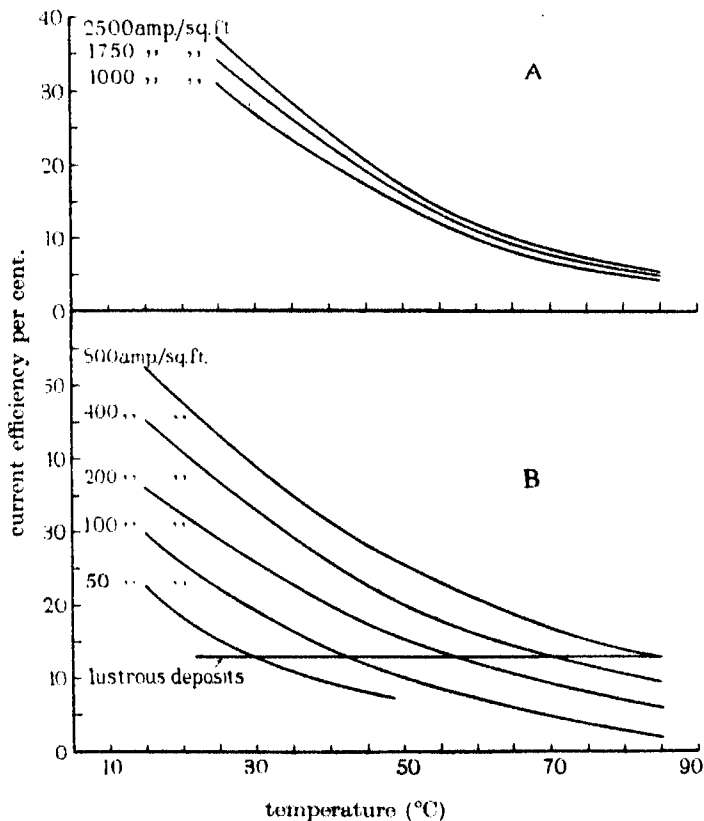
FIGURE 7

will be seen that the shapes of the curves in figures 7 A, B are essentially similar. For the three curves we have obtained, it is found that bright deposits are formed at the same efficiency—approximately 8 %—and that the transition from zone I to zone II (figure 4) is characterized at each current density by an efficiency of approximately 15 %. At constant current density, the efficiency decreases steadily with increasing temperature of the electrolyte. At constant temperature, the increase of efficiency with current density follows an exponential curve, i.e. the same percentage increase in efficiency results from increasing the current density from 100 to 200 amp./sq.ft. as is caused by an increase from 1000 to 2000 amp./sq.ft. It appears, further, that

similar types of deposit, e.g. lustrous, are produced at the same deposition efficiencies from the same solution.

REFERENCES

Arkharow, W. 1936 *Tech. Physics U.S.S.R.* **3**, no. 12, 1072.
Biriükoff, N. D. 1936 *Korrosion u. Metallsch.* **12** (7), 165.
Blum, W. J. 1932 *J. Franklin Inst.* **213**, 17.
Cymboliste, M. 1938 *Trans. Amer. Electrochem. Soc.* **73**, 353.
Frye, J. H. & Hume-Rothery, W. 1942 *Proc. Roy. Soc.* A, **181**, 1.
Glocker, R. & Kaupp, E. 1924 *Z. Phys.* **24**, 121.
Schneidewind, R., Urban, S. F. & Adams, R. C. 1928 *Trans. Amer. Electrochem. Soc.* **53**, 499.
Shishkin, V. & Gernet, H. 1928 *Z. Elektrochem.* **34**, 57.
Stoney, G. G. 1909 *Proc. Roy. Soc.* A, **82**, 172.
Wood, W. A. 1931 *Phil. Mag.* **12**, 853.
Wood, W. A. 1935 *Trans. Faraday Soc.* **31**, 1248.
Wright, L., Hirst, H. & Riley, J. 1935 *Trans. Faraday Soc.* **31**, 1253.

Applications of the photographic method to problems in nuclear physics. I

(a) The determination of the energy of homogeneous groups of α-particles and protons

(b) The determination of the energy of fast neutrons

BY C. F. POWELL, *The H. H. Wills Physical Laboratory, University of Bristol*

A detailed account is given of an application of the photographic method to the determination of the energy of homogeneous groups of particles from nuclear reactions. The particles studied are (a) α-particles from thorium and uranium; (b) protons from the disintegration of boron and lithium by 500 ekV deuterons, and (c) neutrons from the deuteron bombardment of deuterium, beryllium, boron and fluorine. It is shown that the experimental width at half maximum of a homogeneous group of particles is nearly independent of the energy in the range from 1 to 13 eMV, and is equal to 0·3 eMV for protons and 0·5 eMV for α-particles. In modern conditions the photographic method provides a powerful and reliable instrument for nuclear research.

1. INTRODUCTION

Knowledge of the atomic nuclei comes largely from observations involving the detection and the measurement of the energy of the various particles and γ-rays resulting from nuclear interactions. In view of the great technical advances which are likely to occur with increasing knowledge and control of nuclear processes it is important to develop more powerful methods of making such observations. New methods are particularly desirable for measurements with neutrons and γ-rays, since the expansion chamber methods which have hitherto been the most powerful for this purpose are slow and arduous.

Just before the war I gave a preliminary account (Powell *et al.* 1939, 1940) of experiments which strongly suggested that in modern conditions the old photographic method had a useful contribution to make to the solution of this problem, and that it deserved much more serious attention than had hitherto been given to it. In particular, it was shown that in suitable conditions the distribution in range of the tracks produced by a homogeneous group of protons gives a measure of the energy of the group which is not inferior to that obtained with the best applications of the counter and expansion chamber methods. In this and subsequent papers a detailed account is given of the methods developed and of their application to a wide range of problems in nuclear physics. Most of the material is now some years old, its preparation in a form suitable for publication having been much delayed by circumstances connected with the war.

2. HISTORY OF THE METHOD

Since the discovery by Kinoshita (1911) of the photographic action of the α-particles, and the subsequent recognition of the fact that with certain types of photographic plates the tracks of any α-particles which have passed through the emulsion are represented after development by a line of developed grains, several observers, and notably Blau & Wambacher (1932, *et al.*), have employed the method for experiments in nuclear physics. An account of the history of the development of the method and a bibliography has been given recently by Shapiro (1941).

Whilst the potential power of the method has been generally recognized, the scope of its application has, until recently, been very restricted. This has been due to technical difficulties. In work with the α-particles emitted by the natural radioactive elements, it was found that the total number of developed grains in the track of a single α-particle with a range of 5 cm. of standard air was small, of the order of twenty. It followed that the length of a track, defined as the distance between the first and last grains recognized as belonging to the track, did not give a sufficiently precise measure of the energy of the particle which produced it. Thus, it was found that if the lengths of the tracks produced in a suitable emulsion by a large number of α-particles of the same initial energy were plotted as a range distribution, the resulting peak had a large spread; an effect analogous to the straggling of α-particles

in a gas, but more serious. The corresponding uncertainty in the determination of the energy of an individual particle from the length of the track it produced in the emulsion made it impossible to apply the method successfully to the study of individual disintegrations involving the emission of α-particles. The individual tracks from a disintegrating nucleus could be recognized, but the accuracy of the determination both of the energy of the particles and of the spacial orientation of the tracks was greatly inferior to that achieved with the expansion chamber.

These difficulties were found to be less serious in work with protons. It is found that the density of grains along a track depends on the ionization density but varies only slowly with it. Thus in a 'half-tone' emulsion the grain density in a proton track is 80 % of that for an α-particle of the same velocity whereas the ionization in a gas is only 25 %. Further, the range increases with energy much more rapidly for a proton than for an α-particle so that the uncertainty in the range determination due to the spacing of the grains leads to a smaller uncertainty in the energy.

In spite of the more favourable conditions thus provided for work with protons, some experimenters concluded that the remaining difficulties are so serious that the method cannot be applied to the accurate determination of proton and neutron energies. Thus Taylor (1935) examined the possibility of determining the distribution in energy of neutrons passing through an emulsion by measuring the lengths of the directly knocked-on proton tracks, and concluded that the uncertainties did not allow the method to be successfully employed for this purpose. Similar experiments were made by Blau & Wambacher (1932, 1934), but like Taylor they had at their disposal only neutron sources produced by bombarding light elements with the α-particles from the naturally occurring radioactive elements and the strength of the neutron sources was relatively small. Even using thick targets and 'poor geometry', long exposures gave so few tracks in unit area of the emulsion that the work of examining the plates was extremely tedious. The energy distribution studied was complicated and it had not been measured by expansion-chamber methods so that a comparison of the two methods was not possible. In these circumstances it is not surprising that the results were not sufficiently conclusive to demonstrate the advantages and importance of the photographic method and to lead to its widespread adoption.

The great potential power of the photographic method lies in the fact that if the range measurements can be made sufficiently accurate and simple, and if the spacial orientation of a track in the emulsion can be determined, then the photographic plate becomes an instrument equivalent to an expansion chamber with certain great advantages. The most important of these are:

(1) Whereas the expansion chamber detects particles during only about 1/1500 of the working time, the photographic plate is continuously sensitive.

(2) In contrast with the automatic expansion chamber which is elaborate and costly to operate, the photographic plate is extremely simple and cheap.

(3) The determination of the range of a particle from stereoscopic pairs of expansion-chamber photographs depends upon subsidiary measurements, such as

the pressure and the constitution of the gas in the chamber, whereas with the photographic method the measurement depends upon a single determination of a direct permanent record of the track.

The following experiments show that with modern sources of particles and using the most suitable microscopic equipment, these advantages can be fully realized.

3. Photographic and microscopic technique

There are three main aspects of the problem of determining the range of charged particles from the tracks they leave in the developed emulsion. First the emulsion must be suitable in giving a sufficiently high grain density along the tracks without a large background. Ilford 'half-tone' plates have been employed throughout the present work. Emulsions of different thickness up to a maximum of 150μ have been used for different purposes. Plates with emulsions of thickness 50μ and less are developed after exposure by immersion for 25 min. in 1 in 15 'Rodinol' at a temperature of $18°$ C, hardened for 10 min. in 1 % acetic acid, and fixed in acid 'hypo' for periods depending on the thickness of the emulsion. In the case of the emulsions thicker than 50μ it is desirable to maintain the various baths at a constant temperature during the processing of the plates, and during the final washing, in order to prevent 'frilling', i.e. the peeling of the emulsion from the glass. Large plates are less subject to 'frilling' than small, and greater liberties can be taken with them. With emulsions thicker than 50μ, in order that the deeper lying parts of the emulsion may be developed, the development time is increased up to 2 hr., with a proportionate decrease in the dilution of the developer.

The second problem is connected with the examination of the plates and the determination of the length of the tracks. In all the plates which we have examined, hitherto, there is a background of grains produced by chemical fog or otherwise, and these must be distinguished as far as possible from the developed grains in the track. Microscope objectives of high aperture have the advantage of a restricted depth of focus with the result that the visible background of grains is reduced as compared with objectives of lower aperture. Further, if a grain is to be associated with a given track, then it must lie not only in the line of the track as seen in the microscope, i.e. in the line which is the projection of the track in the focal plane of the objective, but also at the depth in the emulsion consistent with it being in the track. The depth of focus of a good oil-immersion objective is sufficiently small to make this second criterion an important aid in deciding whether a grain should be regarded as belonging to a given track and, hence, in deciding the correct 'length' of the track.

The use of oil immersion objectives has the additional advantage over 'air' objectives that the refractive index of the emulsion is sufficiently close to that of immersion oil to ensure that the optical quality of the image of grains lying deep in the emulsion is good even with the thickest emulsions employed (150μ). It seems

probable that emulsions of 250μ thickness could be examined successfully if they could be developed in depth.

Because of these advantages oil-immersion lenses of high aperture have been employed in the present work. They are used in conjunction with high-quality substage condensers with a maximum numerical aperture of 1·30. Some observers have reported that contact with immersion oil is harmful to photographic emulsions. Although I have not found this, in order to protect the emulsions from possible damage, I covered them with a number '0' cover-slip, thickness 0·1 mm., mounted on Canada balsam in xylene. I have had experience with a number of objectives and find the most suitable to be a 3 mm. apochromat, numerical aperture 1·30, magnification 60 and free-working distance 0·25 mm. This objective is normally employed in conjunction with a × 6 'Periplanatic' eyepiece provided with an eye-piece scale 5 mm. long divided into 100 parts. With this arrangement and in the 'half-tone' emulsion the range of a particle which appears to equal the length of the scale has been found to be 14·6 cm. of air, its actual length in the emulsion being 0·1 mm. Other eyepieces giving higher magnification are occasionally employed for special purposes.

The third problem is to reduce to a minimum the nervous strain on the observer in counting or measuring large numbers of tracks. Experience has shown that in this connexion it is a great advantage to use binocular microscopes with inclined eyepieces. Even small improvements in the optical quality of the objectives, leading to improved definition of the image, gives appreciable relief to the observer, and the apochromat of which the characteristics are described above appears to be the best available. Apochromats with an aperture of 1·4 are obtainable, but their free working distance is too small for them to be used for the examination of the thickest emulsions, if a cover slip is used.

DETERMINATION OF THE ENERGY OF CHARGED PARTICLES

(4) α-particles from thorium and uranium

The first experiments were made on the tracks of thorium α-particles, obtained by soaking an emulsion in a weak solution of thorium nitrate, drying, leaving it for some days and subsequently washing and processing the plate. It is well known that an emulsion so treated shows the presence of characteristic thorium stars: groups of α-particles radiating from a single centre, and corresponding to the successive disintegration of thorium nuclei scattered throughout the emulsion. Some of the tracks lie nearly in the focal plane of the objective and their length can be measured directly to the nearest eyepiece scale division. The lengths of 750 tracks were measured and the corresponding frequency-range curve constructed. There was a tendency for the observer to measure the lengths to the nearest even digit and, to eliminate this bias, the number of tracks in three successive values of the range in eyepiece scale divisions were summed and plotted against the mean range as shown in figure 1a.

It will be seen from figure 1*a* that five groups of α-particles are clearly resolved. The ranges of the particles from the complete succession of disintegrations of thorium are shown by arrows in the same diagram, the stopping power of the emulsion being determined from the experiments with protons described later. The two ranges of 4·8 and 5·0 cm. of air are unresolved but the two groups together produce a broader and higher group than those due to the single groups of particles. No attempt was made to ensure that a fair sample of the tracks was taken for measurement, and no significance is to be attached to the fact that the peak of lowest range is less intense than the others.
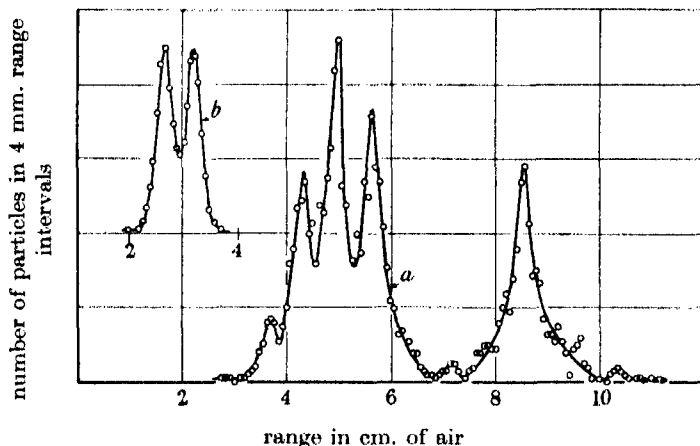
FIGURE 1. Curve (*a*), distribution in range of the α-particles arising from the successive disintegrations of thorium. Curve (*b*), distribution in range of the α-particles from uranium Ur²³⁴ and Ur²³⁸.

An examination of figure 1 shows that whilst the peaks present very sharp maxima, so that the energy of homogeneous groups of α-particles can be deduced from such observations with considerable accuracy, the 'wings' of the distribution are relatively broad, so that individual determinations of particle energy from the lengths of single tracks are subject to relatively large errors.

Figure 1*b* shows the distribution of the lengths of the tracks of α-particles from uranium obtained by similar methods, the two groups arising from Ur²³⁴ and Ur²³⁸. Here again no precautions were taken to obtain a true sample of the two groups and no significance can be attached to the relative intensities of the two groups.*

(5) *Protons from the transmutation of boron and lithium by deuterons*

For the examination of the accuracy of the method for work with protons experiments were made with the protons resulting from the disintegration of boron and lithium under bombardment with 500 kV deuterons provided by the Bristol 700 kV generator, the target system being shown in figure 2. The disintegration protons

---

* *Footnote added in proof.* Reproductions of murophotographs of these tracks have been published recently—Powell, 1942.

emerged from the vacuum chamber through a mica window, and then passed into the plate holder through a thin aluminium foil to enter the emulsion at a glancing angle of 5°. The plate holder contained air at atmospheric pressure, and in examining the plates, allowance was therefore made for the fact that the total stopping power of material through which the protons passed before entering the emulsion was different for different areas of the emulsion.
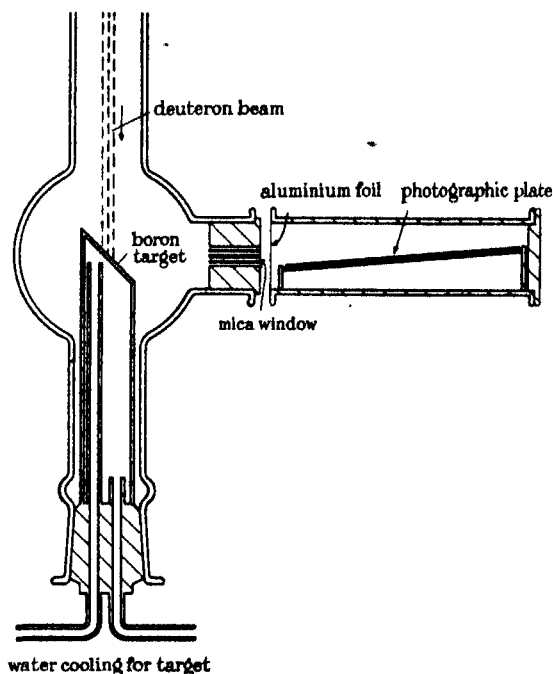
FIGURE 2

Owing to the presence of elements of high nuclear charge in the emulsion, such as silver and bromine, the proton tracks are more subject to scattering than is the case when using the hydrogen or air-filled expansion chamber, or with counter methods of detection, but except in rare instances and at the ends of the tracks the angles of scattering are less than 5°. A track can be easily recognized if it makes an angle of 'dip' into the emulsion of less than 25° with respect to the focal plane of the objective.

If therefore we define the range in the emulsion as the quantity $\int ds$ measured along the length of the track, we can expect the errors introduced into the measurements by elastic scattering of the protons in the emulsion to be negligible. With protons of high energy we must expect inelastic collisions with the nuclei of the atoms forming the emulsion to take place, collisions which leave the struck nuclei in excited states and from which the proton leaves with reduced energy. Experiments (Wilkins 1940, 1941; Powell et al. 1940) show, however, that such collisions are very rare com-

pared with elastic collisions, and it is therefore certain that no appreciable error will be introduced into the measurements by processes of this kind.

In the case of tracks whose images are longer than the eyepiece scale, the length is determined by first setting the scale parallel to the initial direction of the track and the beginning of the scale coincident with the first grain in the track. A grain in the emulsion is chosen which coincides with the end of the scale, irrespective of whether it lies in the track or not, and the mechanical stage is moved until this grain coincides with the other end of the scale. This operation is continued until the end of the track lies within the limits of the scale. If the direction of the track changes at any point corresponding to the scattering of the original particle, the scale is rotated to lie parallel to the new direction.
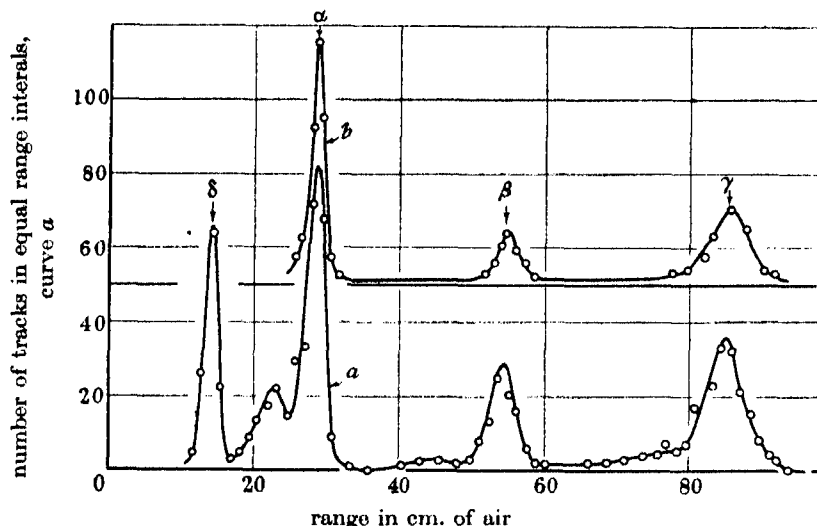
FIGURE 3. Distribution in range of the protons from the transmutation of boron by 500 ckV deuterons. Curve (a), photographic method. Curve (b), counter experiments of Cockcroft & Lewis (1936).

The frequency-length curve, deduced from measurements on 700 proton tracks from the disintegration of boron, is shown in figure 3a. The range distribution of protons from the same reaction obtained by counting methods as determined by Cockcroft & Lewis (1936) is shown in figure 3b, the range scale and origin in the two curves having been adjusted to make the peaks 'α' and 'γ' in the two curves to have the same separation, and the corresponding peaks to come one above the other. The results obtained by the photographic method are clearly satisfactory in that the general form of the proton-range distribution obtained is closely similar to that obtained by counting methods, and that the 'straggling' of the individual peaks is similar in the two cases.

From the results represented in figure 3 it was possible to determine the stopping power of our emulsion since the distance between peaks 'α' and 'γ' is known both in terms of cm. of air from the counter experiments and in terms of our microscope

eyepiece scale divisions. The value obtained is that given in § 3. With this calibration the range in cm. of air of peak 'δ' can be calculated and is found to be 14·6 cm. of air. This is equal within the errors of experiment to the expected range of the protons from the reaction $D^2 + D^2 \rightarrow H^3 + H^1$ when the bombarding deuterons have an energy of 500 kV and the peak is evidently due to the contamination of the target with heavy hydrogen. Figure 4 shows the experimental results transformed to a distribution in energy, and it will be seen that the width of the peaks at half maximum is nearly constant in the range from 2·8 to 8·5 eMV and equal to ~0·3 eMV. It follows that with protons in this range of energy, the uncertainties due to straggling are not worse using the photographic method than those met in experiments with counters.
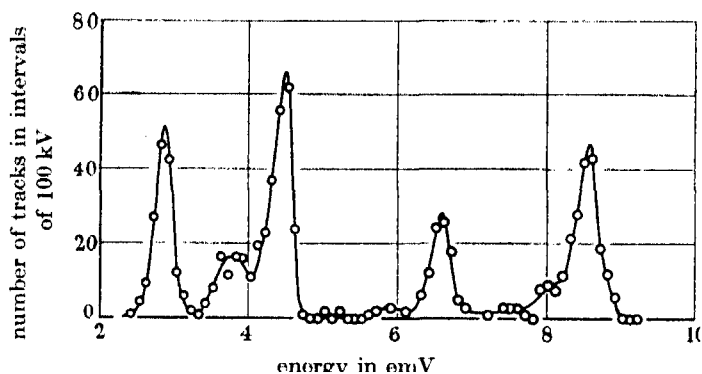
FIGURE 4. Distribution in energy of the protons from the transmutation of boron by deuterons as obtained by the photographic method. The peak at lowest energy is due to the presence of deuterium absorbed in the target. The width of the different peaks at half maximum is less than 0·3 eMV in each case.

The distribution in range of the particles resulting from the transmutation of lithium by 500 ekV deuterons has been investigated by similar methods and the results are represented in figure 5. Curve (a) gives the distribution as determined in the present experiments, and curve (b) shows the corresponding results obtained by Rumbaugh & Hafsted (1936) using counter methods of detection. The particles of longest range are protons from the reaction $Li^6 (d, p) Li^7$, the Li^7 being formed in the ground state or in an excited state at 0·44 eMV. The low range peak is due to α-particles from the reaction $Li^6 (d, \alpha) He^4$, and the particles of lowest range constitute the high-energy end of the continuous distribution of α-particles from the reaction $Li^7 + H^2 \rightarrow 2He^4 + n^1$.

As in the experiments with the boron protons, the range scale of the results obtained by the photographic method has been adjusted to make the high- and low-range peaks fall under the corresponding peaks as determined by the counter experiments. We are thus provided with a second determination of the stopping power of the emulsion, and the value so obtained agrees with that previously determined with an error of less than 1 %. Figure 5 also shows the results transformed to an energy scale in the case of the α-particle and proton groups, the width of the

peaks at half maximum being 0·3 and 0·5 eMV for protons and α-particles respectively.

The observed width of the α-particle peak in figure 5 is greater than that obtained in the observations on the thorium α-particles. This is due to the fact that with the thorium stars the beginning of any track is subject to very small uncertainties, since we take it to be the point of origin of a number of tracks radiating from a common centre. With the lithium α-particles, however, we take the beginning of a track to coincide with the first developed grain recognized as belonging to the track. The range determinations of the thorium α-particles are therefore subject to roughly half the error of the lithium α-particles, and this is reflected in the increased energy spread of the latter.
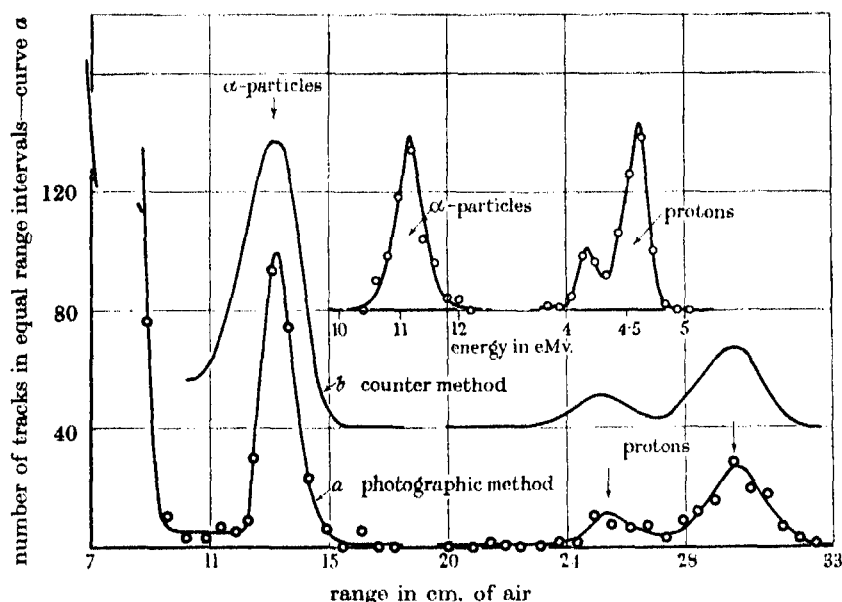
FIGURE 5. Particles from the transmutation of lithium by deuterons. Curve (*a*), range distribution by the photographic method. Curve (*b*), range distribution from counter experiments of Rumbaugh & Hafstad (1936). Curves (*c*) and (*d*), energy distribution of the α-particles and protons from the photographic measurements.

In the present experiments the groups of particles from the various transmutations are inhomogeneous as a consequence of the penetration of the thick target by the bombarding deuterons. We may therefore infer on the basis of the present results that using the photographic method the experimental width at half maximum of a homogeneous group of protons, in the energy range from 1 to 9 eMV, is less than 0·25 eMV; and of α-particles, less than 0·5 eMV. The intensity of the different peaks are also in satisfactory agreement with the values obtained by the counter methods. In this respect the photographic method has the particular advantage that particles of different range are simultaneously recorded in the emulsion so that determinations

of the relative intensity of different peaks are independent of fluctuation in the intensity of the bombarding stream.

We may conclude that the photographic method provides a reliable method for determining the energy and intensity of homogeneous groups of protons and α-particles. This suggested that the method might be used with great advantage for scattering experiments using the high-energy proton and deuteron beams provided by the cyclotron, and this application has now been made. It also showed that it was reasonable to attempt to apply the method to the determination of the energy of groups of fast neutrons.

DETERMINATION OF THE ENERGY OF FAST NEUTRONS

(6) *Neutrons from beryllium and boron under deuteron bombardment*

The most satisfactory method hitherto developed for determining neutron energies in the range of 1 eMV and upwards is that based upon the hydrogen- or helium-filled expansion chamber. If a neutron of energy E_0 collides with a proton at rest and throws it forward in a direction making an angle θ with the original direction of motion of the neutron, then the proton receives an energy $E_0 \cos^2\theta$.* If $\theta \leqslant 5°$, $\cos^2\theta \sim 0.99$ and the initial energy of the proton after impact is sensibly equal to the original energy of the neutron. If high-energy neutrons are passed through the plate, then, since a photographic emulsion contains combined hydrogen, one may expect to find the tracks of knocked-on protons in it after development. If the direction of the neutrons passing through the plate is known, then one can choose those protons which make only a small angle with the direction of motion of the neutrons, measure their ranges and deduce the distribution in energy of the neutrons forming the original directed stream.

The first experiments were made with the Bristol generator on the neutrons resulting from the disintegration of the light elements under deuteron bombardment and the apparatus is shown in figure 6. It is designed to reduce to a minimum the amount of material in the neighbourhood of the neutron source in order to reduce the scattering of the neutron beam. The deuteron beam passes through the hole in the quartz disk, 4 mm. in diameter, and is focused on to the target. When the beam passes directly through the middle of the hole in the quartz disk its edges become just red hot under the impact of the particles in the periphery of the beam. If the beam becomes displaced, then the glow from the quartz becomes unsymmetrical and, if the main beam hits it, glows white hot and rapidly disintegrates at the point of impact. The quartz thus provides an indicator which enables us to maintain the beam focused on the centre of the target. Focusing is carried out by voltage control of the first accelerating gap and can be made extremely sharp. The greater part of the beam strikes an area of the target about 2 mm. in diameter and the neutron source is thus highly localized. The target system is cooled by water, the whole construction

* It is assumed that radiation losses, neutrino interactions, etc., are negligible.

being kept as small and thin as possible.* The neutrons from boron and beryllium were first examined because the energy spectrum of these neutrons had already been examined by the expansion-chamber methods and would therefore provide a reliable test of the accuracy of the method.

Twelve plates, each measuring 5 by 2 cm., were arranged in small metal boxes around the neutron source in a position, as shown in figure 6, to detect neutrons emitted at an angle making about 100° with the direction of the bombarding stream, the neutrons passing through the emulsion in a direction parallel to its surface. The plates were given different exposures so that one among them would contain a convenient number of tracks per field of view of the microscope for examination and measuring. With the middle of the plate at a distance of about 5 cm. from the neutron source and with 30 μA in the deuteron beam (unresolved) exposures of about 2 min. with a lithium target, 15 min. with a beryllium target, and 40 min. with a boron target were found to be satisfactory.

After development the plates were mounted on the mechanical stage of a microscope, provided with 'x' and 'y' motion, in such a way that the original direction of the neutron beam was parallel to one of the directions of motion of the mechanical stage. Only those tracks were measured which made small angles of dip from the plane of the emulsion and which were inclined at angles less than 5° to the direction of the neutron beam through the plate. The angle which the projection of the track makes with the direction of the neutron beam can be measured by providing one of the eyepieces with a fiducial line, which in its normal position is kept parallel to the direction of the neutrons in the plate, and measuring the angle through which this eyepiece must be turned in order to make the line coincide with the track. The angle of dip can be estimated with sufficient accuracy by taking those tracks which are seen in focus over a certain length of the track. With a little experience tracks which make an angle of less than 5° with the original direction of the neutron beam

deuteron beam

window

quartz disc

water cooling

photographic plate

target

water cooling
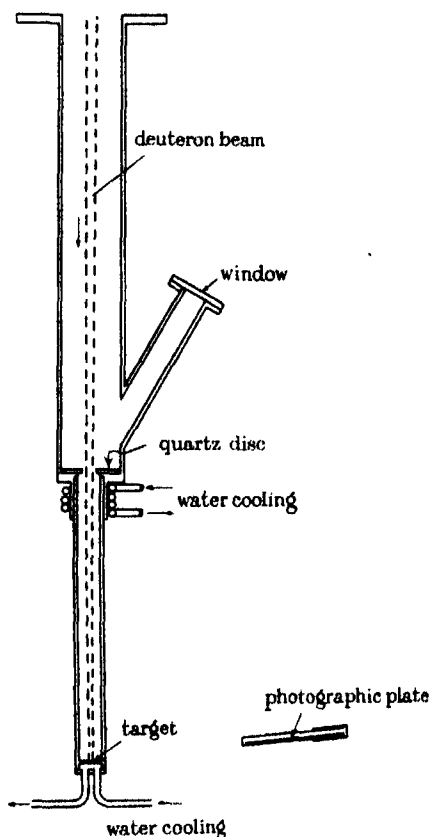
FIGURE 6. Apparatus for the investigation of the neutrons from the light elements under deuteron bombardment.

---

* Subsequent experiments on the elastic and inelastic scattering of neutrons have shown that elaborate precautions of this type are unnecessary.

can be quickly identified without recourse to actual measurement of the precise angle by the methods described in a later paper.

In order to obtain a true sample of the tracks which satisfy the angular requirements, and to avoid errors due to any tendency on the part of the observer to choose either the longer or the shorter tracks, it is sufficient to measure all tracks within a given area which satisfy the angular requirements. As a further precaution against prejudice the plates were examined independently by two observers using microscopes with different magnifications. The two results were then transformed to the same range scale and from this result the energy distribution deduced.
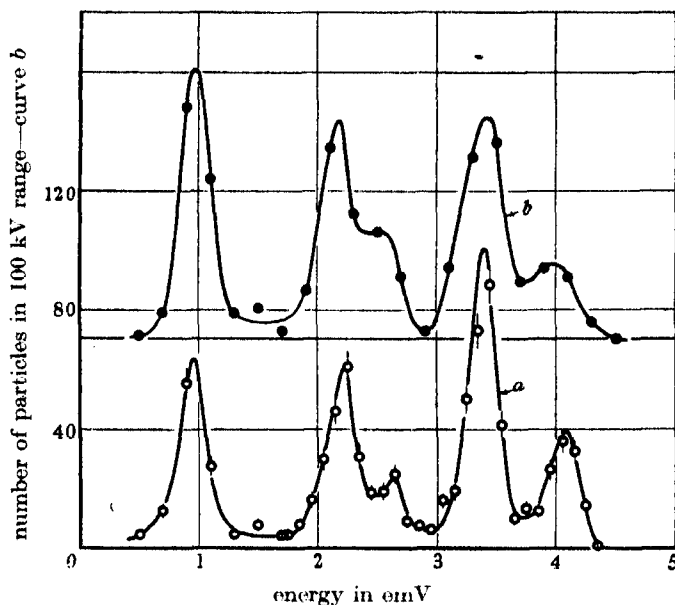
FIGURE 7. Distribution in energy of the protons projected by the neutrons from beryllium bombarded by 500 ekV deuterons. Curve (a), photographic method. Curve (b), expansion chamber results of Staub & Stephens (1939).

The results obtained from the bombardment of beryllium and boron targets are shown in figures 7 and 8, the expansion-chamber measurements being reproduced for comparison in the case of beryllium. It will be seen from figure 7 that the width at half maximum of the different groups is ~ 0.3 eMV, as with the proton groups previously examined. In the case of boron, the group at 13 eMV, which arises from the reaction $B^{11}(d, n) C^{12}$, the C^{12} being left in the ground state, has a half-width of about 0.5 eMV. The peaks of lower energy, corresponding to the formation of C^{12} in excited states, have a half-width of 0.7 eMV. This is more than twice the value to be expected for a homogeneous group of particles in this range of energies and must be due to a natural width in the excited levels of C^{12} or to the fact that they consist of unresolved multiplets.

The measurements demonstrate clearly the power of the method in dealing with

neutron experiments. The spectrum of the neutrons from each element was deduced from an examination of about 2 cm.² of a single photographic plate. The material for the measurement of a thousand times as many tracks could have been obtained without difficulty during the same exposure. The speed with which results can be obtained depends to some extent on the experience of the observer, but it is not difficult to measure 200 tracks a day. With three observers it is therefore possible to make an analysis of a complicated neutron spectrum in three days. This time is to be compared with a period of from three to six months which is necessary to obtain the same information by expansion-chamber methods. Thus in investigating the
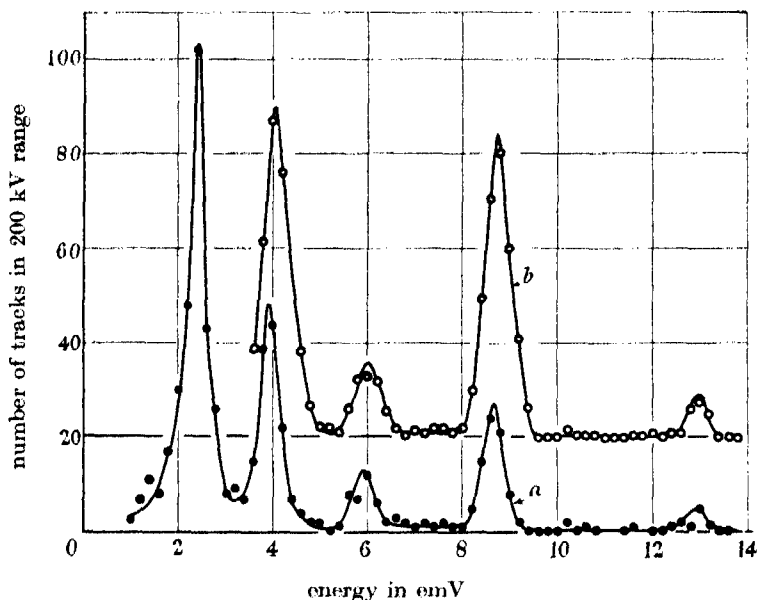
FIGURE 8. Distribution in energy of the protons projected by the neutrons from the transmutation of boron by 500 ekV deuterons, by photographic method. Curve (*a*), 700 tracks. Curve (*b*), measurements on 600 tracks with bias in favour of choosing tracks of long range to increase the accuracy of the distribution in the high energy region.

neutrons from beryllium under deuteron bombardment, Staub & Stephens made 20,000 stereoscopic pairs of expansion-chamber photographs from which 3000 tracks were chosen as suitable for measurement. Apart from the great saving in money and labour which the method thus affords, it is evident that experiments with very weak neutron beams, which could not be attempted with the expansion chamber, become possible. Such effects can be investigated more rapidly by other methods, such as those based upon the radioactivity induced in certain nuclei by neutrons above a certain critical energy, but in many cases the photographic method has the advantage that the information which it provides is much more detailed than that obtained by other methods.
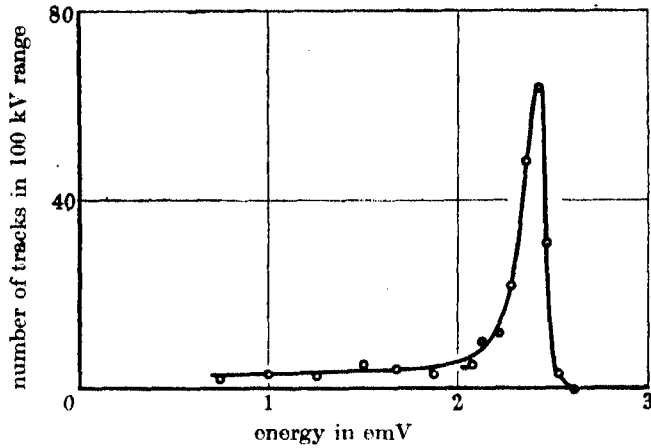
Fɪɢᴜʀᴇ 9. Distribution in energy of the protons projected by the neutrons from the transmutation of deuterium by deuterons.
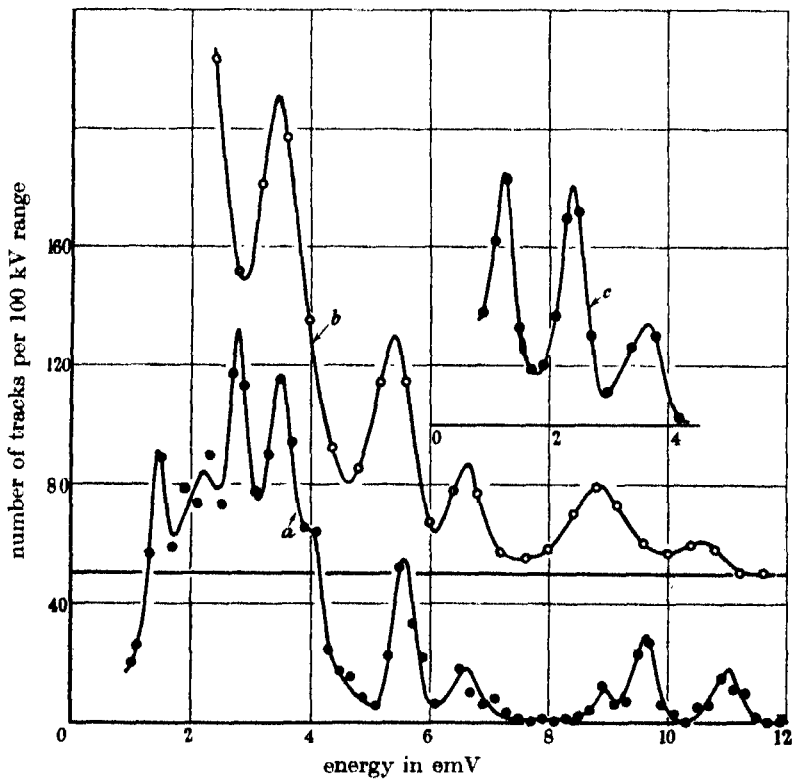
FIGURE 10. Curve (a), distribution in energy of the protons projected by the neutrons from the transmutation of fluorine by 900 ekV deuterons, as determined by the photographic method. Curve (b), distribution in energy of the neutrons from the transmutation of fluorine by 900 ekV deuterons, from the expansion chamber experiments of Bonner (1940).

(7) *Neutrons from the transmutation of deuterium, and fluorine by deuterons*

Two other neutron spectra have been examined by the same method. In figure 9 is shown the analysis of the neutrons from a boron target bombarded by low-energy deuterons. They are from the reaction $D_1^2 + D_1^2 \rightarrow He_2^3 + n_0^1$, the deuterium being adsorbed in the surface layers of the target. Like other observers we find no evidence for a group of low-energy neutrons from this reaction due to the existence of a postulated excited state in He_3. Figure 10 shows the results obtained by bombarding fluorine with 900 kV deuterons from the Cambridge high-tension set. The last result was obtained through the kindness of Professor Cockcroft and Mr Dee, who made it possible for me to make the necessary exposures. This spectrum has also been examined by Bonner (1940) by the expansion-chamber method. His results were not published when the present measurements were made, so that there was no possibility in this case of the analysis by the photographic method being influenced by the results of previous experiments. Once again the agreement between the two methods is satisfactory. In addition to the tracks represented in figure 10, two tracks were observed with energies of 13·2 and 13·4 eMV respectively which are certainly due to boron contamination of the target. Such contamination may account for the peaks of low intensity at 9·3 and 4·3 eMV in figure 9.

The interpretation of this neutron spectrum in terms of the excited states of Ne^{20} has been discussed by Bonner, and further confirmation for the existence of the first excited state at 1·5 eMV has now been obtained in the inelastic scattering of protons by neon (Powell *et al.* 1940).

REFERENCES

Blau & Wambacher 1932 *S.B. Akad. Wiss. Wien,* **141**, 617.
Blau & Wambacher 1934 *S.B. Akad. Wiss. Wien,* **143**, 401.
Blau & Wambacher 1934 *J. Phys. Radium,* **5**, 61.
Bonner 1940 *Proc. Roy. Soc. A,* **174**, 339.
Cockcroft & Lewis 1936 *Proc. Roy. Soc. A,* **154**, 251.
Kinoshita 1910 *Proc. Roy. Soc. A,* **83**, 432.
Powell 1940 *Nature, Lond.,* **145**, 155.
Powell 1942 *Endeavour,* **1**, 151.
Powell & Fertel 1939 *Nature, Lond.,* **144**, 115.
Powell, May, Chadwick & Pickavance 1940 *Nature, Lond.,* **145**, 893.
Rumbaugh & Hafsted 1936 *Phys. Rev.* **50**, 681.
Shapiro 1941 *Rev. Mod. Phys.* **13**, 58.
Staub & Stephens 1939 *Phys. Rev.* **55**, 135.
Taylor 1935 *Proc. Roy. Soc. A,* **150**, 382.
Wilkins 1941 *Phys. Rev.* **60**, 365.
Wilkins & Kuerti 1940 *Phys. Rev.* **57**, 1081, 1082.

The reflexion of X-rays from the 'anti-phase nuclei' of AuCu₃

By A. J. C. WILSON, *Cavendish Laboratory, Cambridge*

Jones & Sykes have observed that the superlattice lines in X-ray photographs of AuCu₃ are not always as sharp as the main lines, and that the broadening depends markedly on the indices of the line. They explain these phenomena by assuming that the crystals of AuCu₃ contain many 'anti-phase nuclei' in which the superlattice is organized in different ways. In the present paper it is shown that the integral breadth of a reflexion from a crystal in which all the unit cells are not the same is $\lambda J_0 / \cos \theta \int J_t dt$, where J_t is the mean value of the product FF^* of the structure factors of two unit cells separated a distance t in the hkl direction. Detailed calculations are made of the broadening to be expected from five different ways in which the nuclei can 'change step'. Closest agreement with the observed broadening is given by a manner of 'changing step' in which the gold atoms avoid one another.

INTRODUCTION

Jones & Sykes (1938) have observed that the superlattice lines in the alloy AuCu₃ are not always as sharp as the main lines, and that the sharpness is a function of the heat treatment. The alloy is face-centred cubic, so that it is possible for the super-lattice to form in four different ways:

A, with gold atoms in the centres of the 100 faces,
B, with gold atoms in the centres of the 010 faces,
C, with gold atoms in the centres of the 001 faces,
D, with gold atoms at the cube corners.
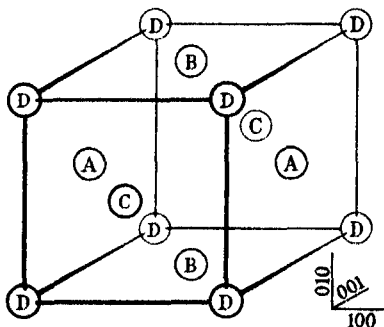
FIGURE 1. The structure of AuCu₃, showing alternative sites for Au atoms.

It appears that within the single crystals practically all the material is ordered, but occasionally it 'changes step', so that the domains of perfect order are smaller than the crystals. The thickness of the transition layer between domains appears to be of the order of one unit cell. Jones & Sykes have worked out the size of the domains on the assumption that they may be treated as 'small particles' and the Scherrer formula applied. The size they obtained depended markedly on the indices

of reflexion. Details of the variation have not previously been published, but
Dr Jones has kindly supplied some typical figures, reproduced in table 1.

TABLE 1. APPARENT PARTICLE SIZES (IN ANGSTROM UNITS) OF THE ANTI-PHASE
DOMAINS IN A SPECIMEN OF $AuCu_3$ (JONES & SYKES, PRIVATE COMMUNICATION)

$h^2 + k^2 + l^2$	apparent particle size					
	film 1	film 2	film 3	film 4	film 5	mean
1	—	—	206	—	192	200
2	96	107	102	—	94	100
5	110	129	120	113	127	120
6	126	149	143	147	146	142
9	122	134	127	149	127	132
10	126	134	113	114	—	122
13	136	218	148	142	129	150
14	102	132	139	129	118	124
17	118	137	143	...	115	128
18	131	166	143	—	126	141
21	105	148	106	—	107	115

The explanation of the variation is to be sought in the manner in which the
changes of step occur. It has been shown (Wilson 1942) that the integral breadths
of reflexions from a structure containing faults of this kind depend on the prob-
ability that two non-adjacent cells are alike. There are several assumptions which
can be made about the way in which the changes of step occur. The simplest is that
they occur without reference to crystallographic direction, i.e. that the probability
of two cells being alike depends only on their distance apart. Another possibility is
that they occur on 100 planes, i.e. the probability of two cells being alike depends on
the projection of their distance apart on the nearest 100 direction. In the present
paper the following possibilities are considered:

(1) mistakes occurring without reference to crystallographic direction, (2) mis-
takes occurring on 100 planes, (3) mistakes occurring on 110 planes, (4) mistakes
occurring on 111 planes, (5) mistakes occurring so that gold atoms do not come into
contact.

It is found that no one of these models gives satisfactory agreement with the
observed apparent sizes. The fifth is the closest approximation, and a combination
of mistakes of the first and fifth kinds is found to give fairly good agreement.

GENERAL CALCULATION†

For the general calculation it is convenient to choose orthorhombic axes such that
the 100 reflexion is that called *hkl* when the crystal is referred to the usual cubic

† The argument in this section is very similar to that of Stokes & Wilson (1942); it is
therefore given in condensed form.

axes. This is always possible. Then the amplitude of the radiation reflected in the direction HKL (H, K and L not necessarily integral) is proportional to

$$\sum_{j_1} \sum_{j_2} \sum_{j_3} F_j \exp\{-2\pi i(j_1 H + j_2 K + j_3 L)\}, \tag{1}$$

where the j's are integers giving the positions of the unit cells of the crystal and the F_j's are the structure factors of the unit cells. The intensity reflected with H between H and $H + dH$, K between K and $K + dK$, L between L and $L + dL$ is proportional to (1) multiplied by its complex conjugate:

$$dI = \sum_{j_1} \sum_{j_2} \sum_{j_3} \sum_{j_1'} \sum_{j_2'} \sum_{j_3'} F_j F_j^* \exp\{2\pi i[(j_1' - j_1)H + (j_2' - j_2)K + (j_3' - j_3)L]\}\,dH\,dK\,dL. \tag{2}$$

The intensity reflected on to the film in the neighbourhood of 100 is

$$\frac{dI}{dH} = \sum_{j_1} \sum_{j_2} \sum_{j_3} \sum_{j_1'} \sum_{j_2'} \sum_{j_3'} F_j F_j^* \exp\{2\pi i(j_1' - j_1)H\}$$

$$\times \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp\{2\pi i(j_2' - j_2)K\}\,dK \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp\{2\pi i(j_3' - j_3)L\}\,dL. \tag{3}$$

The integrals vanish unless $j_2' = j_2$ and $j_3' = j_3$, when they have the value $+1$. Then

$$\frac{dI}{dH} = \sum_{j_1} \sum_{j_2} \sum_{j_3} \sum_{j_1'} F_j F_j^* \exp\{2\pi i(j_1' - j_1)H\}, \tag{4}$$

the total intensity is

$$I = \int_{\frac{1}{2}}^{1\frac{1}{2}} (dI/dH)\,dH = \sum_{j_1} \sum_{j_2} \sum_{j_3} F_j F_j^*, \tag{5}$$

and the intensity at $H = 1$ is

$$\left(\frac{dI}{dH}\right)_0 = \sum_{j_1} \sum_{j_2} \sum_{j_3} \sum_{j_1'} F_{j_1 j_2 j_3} F_{j_1' j_2 j_3}^*. \tag{6}$$

Let a, b, c be the lengths of the orthorhombic axes, let J_t, where $t = ma$, be the mean value of FF^* for two cells separated by a distance ma in the 100 direction (i.e. the hkl direction when referred to the usual cubic axes), and let the volume of the part of the crystal in which the unit cells have others at a distance ma from them be V (the V_x of Stokes & Wilson). Then

$$I = V_0 J_0 / abc, \tag{7}$$

$$\left(\frac{dI}{dH}\right)_0 = \frac{1}{abc} \sum_m J_t V_t$$

$$= (1/a^2 bc) \int J_t V_t\,dt, \tag{8}$$

on replacing the summation by an integration. If the integral breadth of a reflexion is defined as its total intensity divided by the intensity for $H = 1$,†

$$\beta = 2I/(dI/d\theta)_0 = 2I/(dI/dH)_0 \, (dH/d\theta)_0$$

$$= \frac{2V_0 J_0/abc}{(1/a^2bc)\,(2a\cos\theta/\lambda)\int V_t J_t \, dt}$$

$$= \frac{\lambda V_0 J_0}{\cos\theta \int V_t J_t \, dt}. \tag{9}$$

If J_t is a constant this is (as it should be) the expression for the integral breadth of a reflexion from a perfect crystal (Stokes & Wilson 1942, equation (17)). In the applications considered in the present paper it happens to have the form

$$J_0 \exp\left(-\delta \,|\, t \,|\right),$$

where δ is a positive quantity large compared with the reciprocal of the dimensions of the crystal. V_t will normally be of the form $V_0 -$ (a polynomial in t), and J_t will become negligibly small before V_t differs appreciably from V_0. To a sufficient approximation therefore equation (9) may be written

$$\beta = \lambda J_0/\cos\theta \int_{-\infty}^{\infty} J_t \, dt \tag{10}$$

$$= \lambda/\cos\theta \int_{-\infty}^{\infty} \exp\left(-\delta \,|\, t \,|\right) dt = \delta\lambda/2\cos\theta. \tag{11}$$

The apparent particle size is defined by Jones (1938) as $\epsilon = \lambda/\beta\cos\theta$. It is perhaps worth remarking that for perfect crystals this is the volume average of the thickness of the crystal measured in the *hkl* direction (Stokes & Wilson 1942, equation (2)). For crystals containing nuclei of the type considered here it is

$$\epsilon = \frac{\lambda}{\cos\theta}\,\frac{2\cos\theta}{\delta\lambda} = \frac{2}{\delta}. \tag{12}$$

CALCULATIONS FOR PARTICULAR MODELS

It is now necessary to calculate J_t for the particular models listed above. The first four are similar and may be considered together. Suppose that for every increase in distance $a\,ds$ measured in a certain direction there is a probability $\alpha\,ds$ of a change of step occurring, where α is a constant and a is the edge of the (cubic) unit cell.

† Under exceptional circumstances it is possible for the maximum intensity of reflexion to occur for non-integral values of the indices. A hypothetical example is hexagonal cobalt containing so many 'mistakes' that its structure is approaching face-centred cubic (Wilson 1942, equation (16)). The 2 comes into the expression for β because it has become conventional to measure integral breadths in $\chi = 2\theta$.

Let the probability that a cell at a distance as from a given unit cell is like the given unit cell be $P(s)$. If no change of step occurs, and the cell at as is like the given cell, the cell at $a(s+ds)$ will also be like the given unit cell. If a change of step occurs, and the cell at as is not like the given cell, the cell at $a(s+ds)$ has a chance of $1/3$ of being like the given cell, since the change may be to any of the three types of cell which the cell at as is not. Then the probability that the cell at $a(s+ds)$ is like the given cell is

$$P(s+ds) = P(s) + P'(s)\,ds = P(s)\,(1-\alpha\,ds) + \{1 - P(s)\}\,\alpha\,ds/3,$$

$$P'(s) + 4\alpha P(s)/3 = \alpha/3, \tag{13}$$

$$P(s) = 1/4 + K\exp(-4\alpha s/3), \tag{14}$$

where K is a constant. Since $P(0) = 1$, K is $3/4$ and

$$P(s) = \{1 + 3\exp(-4\alpha s/3)\}/4. \tag{15}$$

Suppose that the unit cells are separated by m_1 unit cells in the x direction, m_2 in the y direction and m_3 in the z direction. The possibilities are that as is

(1) the distance between the cells,
(2) the distance between the cells measured in a 100 direction,
(3) the distance between the cells measured in a 110 direction, or
(4) the distance between the cells measured in a 111 direction.

Let n_1, n_2, n_3 be the positive values of m_1, m_2, m_3 arranged in decreasing order of magnitude. Then for the four cases,

$$\begin{aligned} &(1)\ \ s = \sqrt{(n_1^2 + n_2^2 + n_3^2)}, \qquad (2)\ \ s = n_1, \\ &(3)\ \ s = (n_1 + n_2)/\sqrt{2}, \qquad\quad (4)\ \ s = (n_1 + n_2 + n_3)/\sqrt{3}. \end{aligned} \tag{16}$$

What it is necessary to calculate, however, is the probability that two cells separated a distance t in the hkl direction should be the same. For this case $m_1 = ht/\sqrt{N}a$, $m_2 = kt/\sqrt{N}a$, $m_3 = lt/\sqrt{N}a$. If u, v, w are the positive values of h, k, l arranged in decreasing order of magnitude, the four values of s are

$$\begin{aligned} &(1)\ \ s = |t|/a, \qquad\qquad\qquad (2)\ \ s = u\,|t|/\sqrt{N}a, \\ &(3)\ \ s = (u+v)\,|t|/\sqrt{(2N)}a, \qquad (4)\ \ s = (u+v+w)\,|t|/\sqrt{(3N)}a. \end{aligned} \tag{17}$$

The possible structure factors are

$$\begin{aligned} F_a &= \mathrm{Cu} + \mathrm{Au}\,\{\exp\pi i(k+l)\} + \mathrm{Cu}\exp\{\pi i(l+h)\} + \mathrm{Cu}\exp\{\pi i(h+k)\}, \\ F_b &= \mathrm{Cu} + \mathrm{Cu}\,\{\exp\pi i(k+l)\} + \mathrm{Au}\exp\{\pi i(l+h)\} + \mathrm{Cu}\exp\{\pi i(h+k)\}, \\ F_c &= \mathrm{Cu} + \mathrm{Cu}\,\{\exp\pi i(k+l)\} + \mathrm{Cu}\exp\{\pi i(l+h)\} + \mathrm{Au}\exp\{\pi i(h+k)\}, \\ F_d &= \mathrm{Au} + \mathrm{Cu}\,\{\exp\pi i(k+l)\} + \mathrm{Cu}\exp\{\pi i(l+h)\} + \mathrm{Cu}\exp\{\pi i(h+k)\}, \end{aligned} \tag{18}$$

where Au and Cu are the structure factors of the respective atoms. The mean value

of the product FF^* for two cells separated a distance t in the hkl direction is therefore

$$
\begin{aligned}
J_t = {} & F_a[P(s)\,F_a^* + \{1 - P(s)\}\,(F_b^* + F_c^* + F_d^*)/3]/4 \\
+ {} & F_b[P(s)\,F_b^* + \{1 - P(s)\}\,(F_a^* + F_c^* + F_d^*)/3]/4 \\
+ {} & F_c[P(s)\,F_c^* + \{1 - P(s)\}\,(F_a^* + F_b^* + F_d^*)/3]/4 \\
+ {} & F_d[P(s)\,F_d^* + \{1 - P(s)\}\,(F_a^* + F_b^* + F_c^*)/3]/4.
\end{aligned}
\tag{19}
$$

If h, k, l are all even or all odd, all the F's are $(Au + 3Cu)$, and J_t is $(Au + 3Cu)^2$, independent of s. The main lines are therefore as sharp as is permitted by the size of the crystals. If h, k, l are not all even or all odd, the F's are $(Au - Cu)$ or $(Cu - Au)$, and (19) reduces to

$$
J_t = (Au - Cu)^2 \exp\left(-4\alpha s/3\right).
\tag{20}
$$

The apparent particle sizes for the four cases are therefore

$$
\begin{array}{ll}
(1)\ \ 3a/2\alpha, & (2)\ \ 3\sqrt{N}\,a/2\alpha u, \\
(3)\ \ 3\sqrt{(2N)}\,a/2\alpha(u+v), & (4)\ \ 3\sqrt{(3N)}\,a/2\alpha(u+v+w).
\end{array}
\tag{21}
$$

Perhaps the easiest way of comparing the predictions from the four models with the observed apparent particle sizes is to arrange the reflexions in order of decreasing apparent size, and give also the predicted orders of apparent size. This is done in table 2. The 'observed' particle sizes for reflexions 221 and 411 have been calculated from the observed sizes from lines 8 and 18 on the assumption that 300 would give the same result as 100, and 330 as 110. Unfortunately line 17 (410+322) cannot be separated in this way. Model 1, which predicts the same apparent size from each line, is omitted. None of the four gives reasonable agreement with the observed order of apparent size. Model 3 is perhaps the closest.

TABLE 2. COMPARISON OF OBSERVED AND CALCULATED APPARENT SIZES

order obs.	$\epsilon_{obs.}$ A	order (2)	order (3)	order (4)	order (5 and 6)	$\epsilon_{calc.}$ (6) A	error %
100	200	221	100	100	100	203	+1·5
411	161	110	411	310	411	157	−2·5
320	150	410+322	410+322	210	320	152	+1·2
211	142	321	211	320	211	135	−5·2
410+322	128	211	310	110	210	131	+9·2
321	124	320	421	410+322	410+322	127	−0·8
310	122	421	221	411	321	122	−1·6
210	120	210	321	421	310	114	−6·6
{421	115	411	210	321	421	112	−2·6
{221	115	310	320	211	221	112	−2·6
110	100	100	110	221	110	109	+9·0

The fifth model, i.e. that in which mistakes occur in such a way that gold atoms do not come into contact, requires more complex treatment. Consideration shows that only on 100 planes is it possible to have such mistakes. In the x direction, for instance, the A arrangement can change to the D, in the y direction to the C, and

in the z direction to the B, without contact of gold atoms. The complete list of possible changes of variety of order is

$$
\begin{array}{ccc}
x \text{ direction} & y \text{ direction} & z \text{ direction} \\
A \to D & A \to C & \left.A \to B\right\rangle \\
B \to C & B \to D & B \to A \\
C \to B & C \to A & C \to D \\
D \to A & D \to B & D \to C
\end{array} \qquad (22)
$$

Suppose that $A_a(x, y, z)$ is the probability that a cell with a displacement ax, ay, az from a given A cell is also A, $B_a(x, y, z)$ the probability that it is B, $C_a(x, y, z)$ the probability that it is C, and $D_a(x, y, z)$ the probability that it is D. Then the probability that a cell at $x + dx, y + dy, z + dz$ is A is

$$
\begin{aligned}
A_a(x + dx, \ldots) &= A_a + (\partial A_a/\partial x)\, dx + (\partial A_a/\partial y)\, dy + (\partial A_a/\partial z)\, dz \\
&= A_a\{1 - \alpha(dx + dy + dz)\} + \alpha D_a dx + \alpha C_a dy + \alpha B_a dz.
\end{aligned} \qquad (23)
$$

The differential equations satisfied by A_a are therefore

$$
\partial A_a/\partial x = \alpha(D_a - A_a), \quad \partial A_a/\partial y = \alpha(C_a - A_a), \quad \partial A_a/\partial z = \alpha(B_a - A_a). \qquad (24)
$$

Similarly the differential equations for B_a, C_a, D_a are

$$
\left.
\begin{aligned}
\partial B_a/\partial x &= \alpha(C_a - B_a), & \partial B_a/\partial y &= \alpha(D_a - B_a), & \partial B_a/\partial z &= \alpha(A_a - B_a), \\
\partial C_a/\partial x &= \alpha(B_a - C_a), & \partial C_a/\partial y &= \alpha(A_a - C_a), & \partial C_a/\partial z &= \alpha(D_a - C_a), \\
\partial D_a/\partial x &= \alpha(A_a - D_a), & \partial D_a/\partial y &= \alpha(B_a - D_a), & \partial D_a/\partial z &= \alpha(C_a - D_a).
\end{aligned}
\right\} \qquad (25)
$$

There are also similar sets of equations giving the probabilities $A_b, B_b, \ldots, A_c, B_c, \ldots$ of the cell at x, y, z being $A, B, \ldots$ when the given cell is $B, C, \ldots$. The solutions which satisfy the boundary conditions $X_x(0, 0, 0) = 1$, $X_y(0, 0, 0) = 0$, are

$$
\left.
\begin{aligned}
A_a &= B_b = C_c = D_d = [1 + \exp\{-2\alpha(y + z)\} + \exp\{-2\alpha(z + x)\} + \exp\{-2\alpha(x + y)\}]/4, \\
A_b &= B_a = C_d = D_c = [1 - \exp\{-2\alpha(y + z)\} - \exp\{-2\alpha(z + x)\} + \exp\{-2\alpha(x + y)\}]/4, \\
A_c &= B_d = C_a = D_b = [1 - \exp\{-2\alpha(y + z)\} + \exp\{-2\alpha(z + x)\} - \exp\{-2\alpha(x + y)\}]/4, \\
A_d &= B_c = C_b = D_a = [1 + \exp\{-2\alpha(y + z)\} - \exp\{-2\alpha(z + x)\} - \exp\{-2\alpha(x + y)\}]/4.
\end{aligned}
\right\} \qquad (26)
$$

In deriving these equations it has been tacitly assumed that x, y, z are positive. Actually all the exponentials should be written with $|x|, |y|, |z|$ in place of x, y, z. The mean value of FF^* for two cells separated by the displacement x, y, z is then

$$
\left.
\begin{aligned}
F_a\{F_a^* A_a + F_b^* B_a + F_c^* C_a + F_d^* D_a\}/4 \\
+ F_b\{F_a^* A_b + F_b^* B_b + F_c^* C_b + F_d^* D_b\}/4 \\
+ F_c\{F_a^* A_c + F_b^* B_c + F_c^* C_c + F_d^* D_c\}/4 \\
+ F_d\{F_a^* A_d + F_b^* B_d + F_c^* C_d + F_d^* D_d\}/4
\end{aligned}
\right\} \qquad (27)
$$

$$
\left.
\begin{aligned}
&= A_a(F_a F_a^* + F_b F_b^* + F_c F_c^* + F_d F_d^*)/16 \\
&\quad + A_b(F_a F_b^* + F_b F_a^* + F_c F_d^* + F_d F_c^*)/16 \\
&\quad + A_c(F_a F_c^* + F_b F_d^* + F_c F_a^* + F_d F_b^*)/16 \\
&\quad + A_d(F_a F_d^* + F_b F_c^* + F_c F_b^* + F_d F_a^*)/16.
\end{aligned}
\right\} \qquad (28)
$$

This takes several forms, depending on the relations between h, k, l:

$$\left.\begin{array}{ll} h, k, l \text{ all even or all odd:} & (Au + 3Cu)^2, \\ h \text{ even or odd, } k \text{ and } l \text{ odd or even:} & (Au - Cu)^2 \exp\{-2\alpha(y+z)\}, \\ k \text{ even or odd, } l \text{ and } h \text{ odd or even:} & (Au - Cu)^2 \exp\{-2\alpha(z+x)\}, \\ l \text{ even or odd, } h \text{ and } k \text{ odd or even:} & (Au - Cu)^2 \exp\{-2\alpha(x+y)\}. \end{array}\right\} \tag{29}$$

The main lines are therefore sharp, as they should be. J_t, the mean value of FF^* for two cells separated a distance t in the hkl direction, is got by replacing x, y, z in these expressions by $|h|t/\sqrt{Na}$, $|k|t/\sqrt{Na}$, $|l|t/\sqrt{Na}$. The apparent particle size from the superlattice lines is therefore

$$\epsilon = \sqrt{(Na)/\alpha(p+q)}, \tag{30}$$

where p and q are the positive values of the indices of the same parity, i.e. the two which are even or the two which are odd. The order of decreasing particle size predicted by this model is given in column 6 of table 2. It will be noted that all lines except 210 are in the correct order. This is much more satisfactory than the previous models, but it is not possible to choose α so as to obtain quantitative agreement. In particular, line 100 should be as sharp as the main lines for all values of α.

This result suggests that a combination of two varieties of mistake, one which occurs only on 100 planes to give a nearly correct order of apparent particle sizes, and another which occurs without reference to crystallographic direction to give a general broadening of the superlattice lines, might produce reasonable quantitative agreement with the observed apparent particle sizes. Suppose that the probability of a mistake of the first kind in a distance $a\,ds$ is $\alpha\,ds$, of the second kind is $\gamma\,ds$. Then, with A_a, etc. having the same meaning as before,

$$A_a + (\partial A_a/\partial x)\,dx + (\partial A_a/\partial y)\,dy + (\partial A_a/\partial z)\,dz$$
$$= A_a\{1 - \alpha(dx + dy + dz) - \gamma(x\,dx/r + y\,dy/r + z\,dz/r)\}$$
$$+ \alpha D_a dx + \alpha C_a dy + \alpha B_a dz + \gamma(1 - A_a)(x\,dx/r + y\,dy/r + z\,dz/r)/3, \tag{31}$$

where $r = \sqrt{(x^2 + y^2 + z^2)}$. The differential equations for A_a are therefore

$$\partial A_a/\partial x = \alpha(D_a - A_a) - 4\gamma x A_a/3r + \gamma x/3r \tag{32}$$

and two similar, with nine more for B_a, C_a, D_a. The solutions are similar to those for the preceding model, but with the exponentials

$$\exp\{-2\alpha(y+z) - 4\gamma r/3\}, \quad \exp\{-2\alpha(z+x) - 4\gamma r/3\}, \quad \exp\{-2\alpha(x+y) - 4\gamma r/3\}.$$

The apparent particle sizes predicted are therefore

$$\epsilon = a/[2\gamma/3 + \alpha(p+q)/\sqrt{N}], \tag{33}$$

where p and q are the positive values of the indices of the same parity. This gives the same order of apparent particle sizes as the previous model, but with

$$2\gamma/3a = 0\cdot00492 \quad \text{and} \quad \alpha/a = 0\cdot00303$$

fair quantitative agreement is obtained. The greatest error (9·2 %) is in line 210, the only one which is in the wrong order.

CRITICISM OF THE THEORY

The agreement between the observed apparent particle sizes and those calculated from model 6 is sufficiently close to give some probability to the idea that even when the nuclei 'change step' the gold atoms tend to avoid one another. At the same time, it does not seem impossible that by the exercise of sufficient ingenuity other models could be found that would give as good or better agreement. The chief objection to the general method of the calculations is that it does not take into account explicitly the 'foam' structure proposed by Bragg (1940).

The author desires to thank Professor Sir Lawrence Bragg, F.R.S., and Dr A. J. Bradley, F.R.S., for their interest in this work.

REFERENCES

Bragg, W. L. 1940 *Proc. Phys. Soc.* **52**, 105.
Jones, F. W. 1938 *Proc. Roy. Soc.* A, **166**, 16.
Jones, F. W. & Sykes, C. 1938 *Proc. Roy. Soc.* A, **166**, 376.
Stokes, A. R. & Wilson, A. J. C. 1942 *Proc. Camb. Phil. Soc.* **38**, 313.
Wilson, A. J. C. 1942 *Proc. Roy. Soc.* A, **180**, 277.

An X-ray study of the dissociation of an alloy of copper, iron and nickel

By VERA DANIEL, PH.D. AND H. LIPSON, M.A., D.Sc.
Cavendish Laboratory, Cambridge

[PLATE 15]

A description is given of the unusual X-ray diffraction effects shown by the alloy Cu_4FeNi_3 while it is in the process of splitting up from one face-centred cubic structure into two. Before the formation of the tetragonal structures described by Bradley a state is formed in which each line of a powder photograph is accompanied by quite strong side-bands. An explanation of this is given and is shown to be in reasonable agreement with most of the quantitative data from the X-ray photographs.

This explanation is that the original cubic lattice is regularly deformed by the segregation of the different atoms. From the positions of the side-bands the directions of the deformations and their average wave-length can be estimated. The intensities, however, do not fit in with the theory; they do not vary exactly in the way predicted. It is shown, however, that the variation is in closer agreement with this sort of modulation of the structure than with any other, but it has not, so far, been found possible to find any factor that will modify the intensities of the side-bands in the way required.

Daniel & Lipson

Proc. Roy. Soc., A, volume 181, plate 15
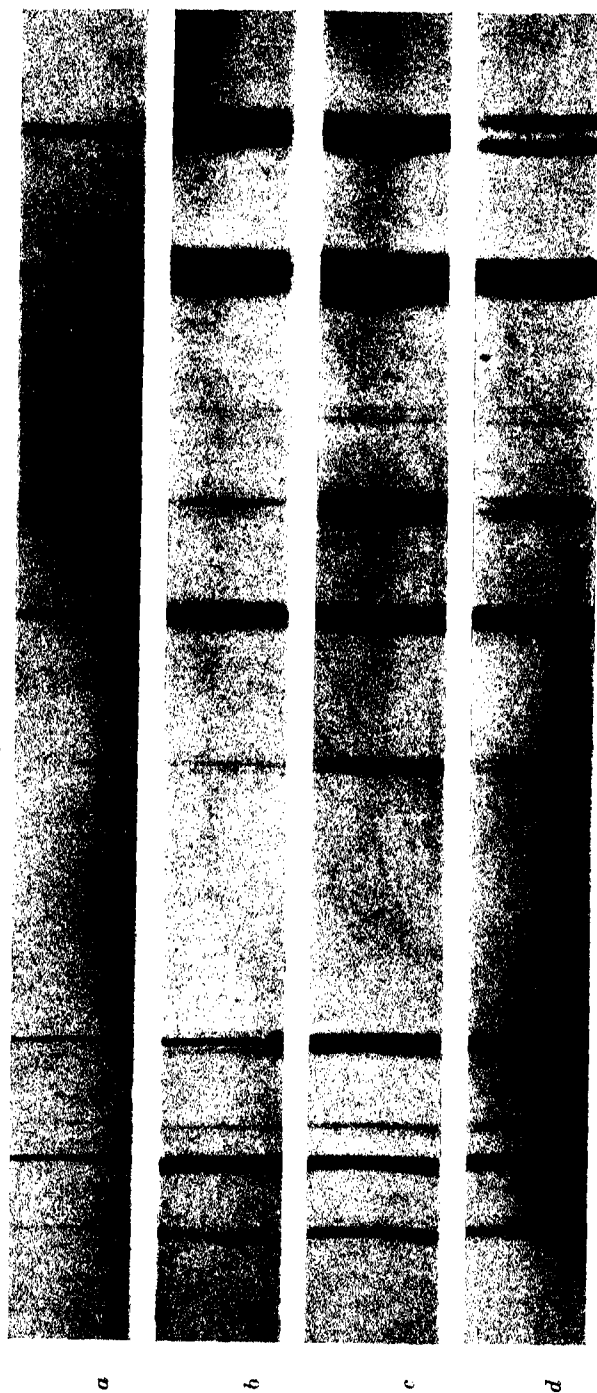
FIGURE 1. X-ray powder photographs of the states in which the alloy Cu$_4$FeNi$_3$ can exist.
(a) Single phase—quenched from 1000° C.
(b) Periodic—(a)+1 hour at 650° C.
(c) Tetragonal—(a)+12 days at 650° C.
(d) Two phase—(a)+11 weeks at 650° C.

INTRODUCTION

In the course of an X-ray study of copper-iron-nickel alloys in this laboratory Dr A. J. Bradley discovered an interesting, and, it is believed, new diffraction effect. The alloy Cu_4FeNi_3 which is single phase (face-centred cubic) above about 800° C, and of which the equilibrium state at lower temperatures consists of two phases of similar structure (Köster & Dannöhl 1935; Bradley, Cox & Goldschmidt 1941) was heat-treated at 650° C for 1 hr. This treatment was not sufficient to destroy the coherence of the crystal lattice; the diffraction lines remained quite sharp; but each one was flanked by two slightly diffuse, but quite strong, side-bands. A photograph is shown in figure 1*b*; this may be compared with the photograph of the single-phase state shown in figure 1*a* (Plate 15).

The alloy is the same as that for which Bradley (1940) has found another metastable state. In this state, each diffraction line is split into a group of lines as shown in figure 1*c*, and Bradley has shown that these groups can be explained by the presence of two tetragonal structures whose axial ratios are nearly unity. The *a* axes were found to be equal and so the natural conclusion to draw was that the two phases fit together on the (100) planes. Comparison of figures 1*b* and 1*c*, however, shows that the new phenomenon is not directly related to the formation of the tetragonal structures, and the fact that each line has a similar structure would suggest that the explanation is fundamentally simpler.

The most obvious suggestion is that the structure is modulated in some way by a periodic variation whose wave-length is much larger than the natural one—the length of the unit cell. The result should then be the formation of side-bands as in wireless theory, and it is necessary to see if such an explanation can fit in with the experimental data.

THEORY

Two sorts of periodic variations can be easily imagined—variations in atomic scattering factor and variations in lattice parameter. We shall consider these in one dimension, as this brings out all the essential points and the mathematics is simpler.

For the first case we may imagine a diffraction grating whose spacing is *a* but of which the scattering power varies sinusoidally with a wave-length Qa. We may put the scattering of the *q*th line as

$$1 + A \sin 2\pi q/Q, \tag{1}$$

where A is a constant less than unity. This may be treated as a grating of spacing Qa with a fine structure. Applying the ordinary crystal-structure formulae (*Internat. Tab.* 1935), the relative amplitudes of the orders of diffraction for such a grating are

$$F_H = \sum_{q=0}^{q=Q-1} [1 + A \sin 2\pi q/Q] \exp(2\pi i Hqa/Qa), \tag{2}$$

where H is the order of diffraction. Thus

$$F_H = \sum \left[1 + \frac{A}{2i} \exp\left(2\pi i q/Q\right) - \frac{A}{2i} \exp\left(- 2\pi i q/Q\right) \right] \exp\left(2\pi i H q/Q\right)$$

$$= \sum \exp\left(2\pi i H q/Q\right) + \sum \frac{A}{2i} \exp\left[2\pi i(H+1)\, q/Q\right] - \sum \frac{A}{2i} \exp\left[2\pi i(H-1)\, q/Q\right]. \quad (3)$$

Each of the summations is a geometrical series and so the sum is

$$F_H = \frac{1 - \exp 2\pi i H}{1 - \exp 2\pi i H/Q} + \frac{A}{2i} \frac{1 - \exp 2\pi i(H+1)}{1 - \exp 2\pi i(H+1)/Q} - \frac{A}{2i} \frac{1 - \exp 2\pi i(H-1)}{1 - \exp 2\pi i(H-1)/Q}. \quad (4)$$

The expression for the intensities $F_H F_H^*$, is obviously rather complicated, but if Q is large it is not necessary to evaluate it. For it can easily be seen that F_H is large only if H, $H+1$, or $H-1$ is a multiple of Q; the relative amplitudes for these three cases are 1, $\frac{1}{2}A$ and $\frac{1}{2}A$.

These orders may then be written as hQ, $hQ+1$, and $hQ-1$, where h is an integer. Since the period of the grating is Q times as long as the spacing of the undeformed grating, it is easily seen that the number h is the index of an order from the undeformed grating. Thus the effect of the distortion we have postulated may be regarded as the production of subsidiary orders $h \pm 1/Q$, of amplitude $\frac{1}{2}A$ relative to the main orders. That is, each diffraction line will be flanked by two weak satellites as shown diagrammatically in figure 2a.
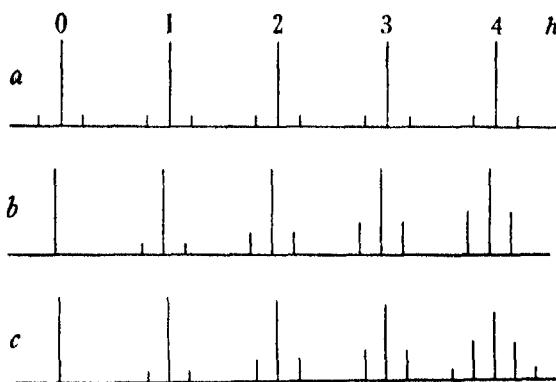
FIGURE 2. Diagrammatic representation of diffraction pattern given by a grating with (a) a periodic variation of the scattering factor; (b) a periodic variation of the spacing, first approximation; (c) a periodic variation of the spacing, rigorous theory.

The second grating to be considered is one in which the positions of the lines vary periodically. This may be expressed by putting the lines at points

$$qa + B \sin 2\pi q/Q, \quad (5)$$

where B is the maximum displacement of any one line from its ideal position. The maximum and minimum distances between the lines obviously occur at the values

of q for which expression (5) is varying most rapidly; at $q = 0$ for instance. Putting $q = 0$ and $q = 1$ in (5) we find that the maximum distance between the lines is

$$a + B \sin 2\pi/Q,$$

which, if Q is large, reduces to

$$a + 2\pi B/Q. \tag{6}$$

Similarly the minimum distance is

$$a - 2\pi B/Q. \tag{7}$$

Since these distances have simpler physical significance than the displacement of any one line from its ideal position, we shall put

$$b = 2\pi B/Q, \tag{8}$$

and b is then the amplitude of variation of the lattice parameter. Thus expression (5) becomes

$$qa + \frac{Qb}{2\pi} \sin 2\pi q/Q. \tag{9}$$

This again may be taken as representing a grating of spacing Qa with a fine structure and the relative amplitudes of the orders of diffraction are

$$F_H = \sum_{q=0}^{q=Q-1} \exp \left\{ 2\pi i\, H\left(qa + \frac{Qb}{2\pi} \sin 2\pi q/Q \right) \middle/ Qa \right\}. \tag{10}$$

If b/a is small enough for Hb/a always to be small compared with unity, second and higher order terms may be neglected in the expansion of $\exp \left\{ i\dfrac{Hb}{a} \sin \dfrac{2\pi q}{Q} \right\}$ and (10) becomes

$$F_H = \sum \exp \left(2\pi i\, Hq/Q \right) \left\{ 1 + i\frac{Hb}{a} \sin \frac{2\pi q}{Q} \right\}. \tag{11}$$

This is similar to expression (2) and may be evaluated similarly. Again the F's are large only for orders hQ, $hQ+1$ and $hQ-1$, and in these cases their relative amplitudes are 1, $hQb/2a$, $hQb/2a$.

A representation of the diffraction pattern is shown in figure $2b$. It will be seen that this differs in an important way from the pattern due to a grating of the first kind: for this each spectrum line had a similar fine structure; for the second grating the fine structure depends on h and the intensities of the satellites are proportional to h^2. Thus, as a special case, the zero order spectrum would have no satellites.

These results are in agreement with those of Dehlinger (1927) and Kochendörfer (1939) who have worked out the rigorous theory of diffraction by a crystal with a sinusoidally distorted lattice. They find that each ordinary diffraction line is replaced by a group of equidistant lines, whose amplitudes are given by $J_n(hQb/a)$ where J_n is the nth order Bessel function. J_0 gives the main line, J_1 the first satellite, etc. Thus the details of the spectra would be as shown by figure $2c$. For $h = 0$, $J_0 = 1$ and all the other J's are zero; thus the zero order has no satellites. If hQb/a

is small, $J_0 = 1$, $J_1 = \frac{1}{2}hQb/a$, and the other orders are negligibly small. This is our approximate case; the justification for introducing it separately is that, as will be shown later, it provides a more convenient basis for comparison with our experimental data.

Other types of grating may exist in which the two types of distortion are combined; Preston (1938) has shown that one result of this may be an asymmetry in the distribution of satellites about the main lines. This may easily be verified by the methods described above.
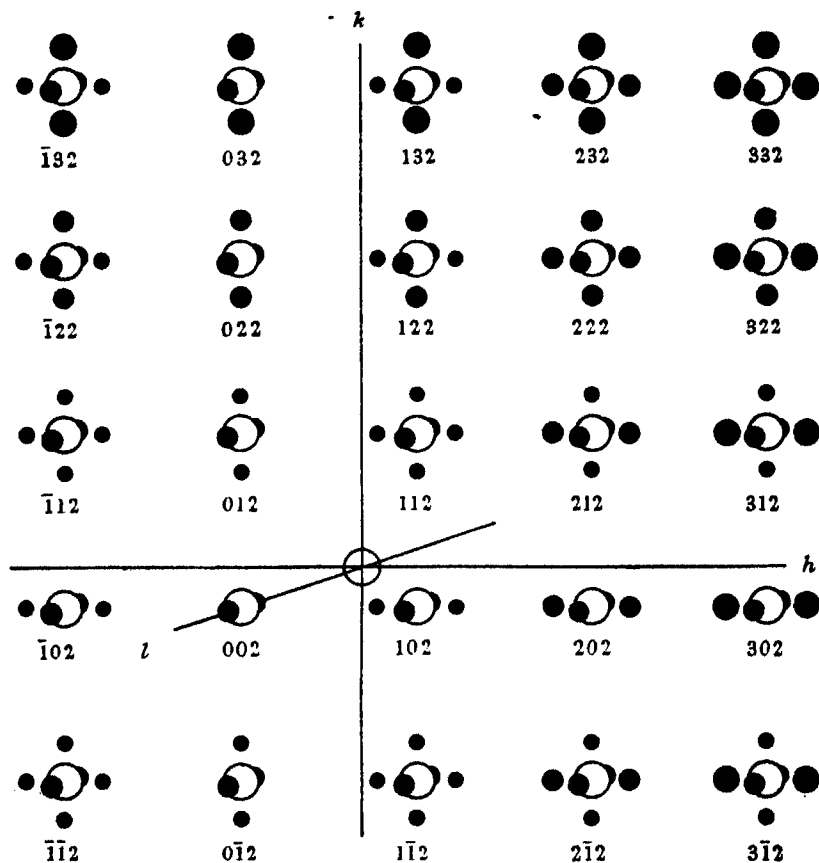
FIGURE 3. Reciprocal points with $l = 2$ for a crystal with three one-dimensional distortions.

It is necessary to adapt our results to the three-dimensional case. Bragg (1941) has pointed out that the satellites produced by a periodic disturbance in a crystal may be simply interpreted in reciprocal space; each reciprocal point will be accompanied by a pair of satellite points, the three lying in a line in the direction of the disturbance and the separation being inversely proportional to its wave-length. In a cubic crystal, however, it is improbable that, without external influence, a disturbance should occur in only one direction; any disturbance in the [100] direction, for instance, is equally likely to occur in the [010] and [001] directions in other parts

of the same crystal. Thus we would have to consider three pairs of spots round each reciprocal point. This, of course, is the simplest possible case; if the disturbance were in any other crystallographic direction the number of satellites would be greater.

The simple case is illustrated in figure 3. The main reciprocal points are represented by circles and the satellites by small black spots. The areas of these spots have been made proportional to the amplitudes that would result from a periodic variation of lattice parameter, according to the results derived from equation (11). Thus in the general case a main point will be accompanied by three pairs of satellites of different intensities, as will be seen, for example, for the point 132 in figure 3. If one index is zero, the corresponding spots will be of zero intensity, and so one would expect only two pairs of satellites. If, however, the periodicity is one of scattering factor each reciprocal point will have the same number of satellite points round it, irrespective of its indices.

Experimental work

The above considerations provide a basis for examining the experimental data, which may conveniently be divided into three categories.

(1) *Numbers of satellite spots*

In order to count these and find their orientations, single crystal photographs are really needed. Attempts were made to obtain such crystals, but they were not successful. Some fragments of alloy with relatively few crystals were obtained, however, and although the orientations of the separate crystals could not be deduced it was verified that there were discrete spots near the main reflexions. Oscillation photographs were taken with CrK, FeK and CuK radiations. None of the reflexions 200, 311 and 400 showed more than two side spots, and none of the reflexions 220 and 331 showed more than four. 222 usually showed four, but in one or two cases there was a suspicion of a greater number; the small scale of the effect, however, made it difficult to observe more than four satellites. Now these numbers of satellite spots are approximately those that would be expected from the reciprocal lattice shown in figure 3; they are right for 200, 220 and 400, and for 311 and 331 a spot corresponding to index unity might be too close in and too weak to be observed. 420 did not show clear satellite spots but this might be explained by the merging together of spots due to indices 2 and 4.

This evidence therefore strongly suggests that the periodicity is of the second type —variation of lattice parameter—and is in the direction of the cube edges. For a complete proof of the second point, however, the orientations of the separate crystals would be required.

(2) *Positions of the side-bands*

Definite confirmatory evidence of the direction of the periodicity can be obtained from powder photographs. Bearing in mind that distances from the origin in reciprocal space are proportional to $\sin\theta$ we can see from figure 3 that the distances

of the side-bands from the main lines will not be simply a function of θ. The separation is a maximum for reflexions that lie in the directions of the distortions and is a minimum for those that are farthest from these directions. In the case considered these are the 100 and 111 reflexions respectively.

Powder photographs were taken with FeK and CuK radiations in a camera of 19 cm. diameter (Bradley, Lipson & Petch 1941); these were microphotometered in order to derive the separation of the side-bands and their intensities. A typical microphotometer curve is shown in figure 4.
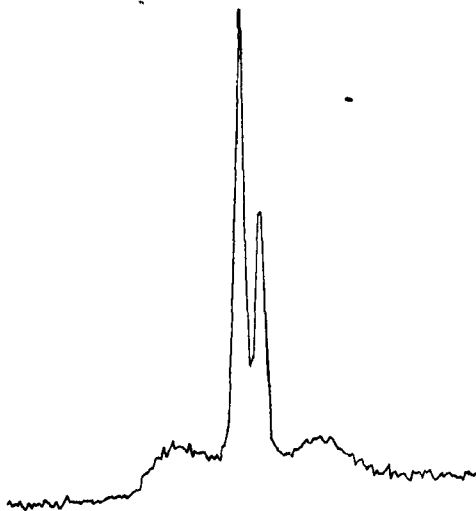
FIGURE 4. Typical microphotometer curve.

It was obvious from the photographs that the separations were roughly correct: the side-bands on line 111 were so close to the main line that they were barely recognizable, while those on 200 were much further out; the separation for line 311 was greater than for line 222 although the latter has a higher value of θ. The measurements supported the visual estimates very well.

To test the theory an expression must be found for the separations of the side-bands in terms of the indices. If θ is the Bragg angle for the reflexion hkl,

$$(h^2 + k^2 + l^2)\, \lambda^2 = 4a^2 \sin^2\theta, \tag{12}$$

where λ is the wave-length of the radiation. Therefore, if the index h changes by a small amount δh while the other two are kept constant, the change in θ, $\delta\theta$, is given by

$$2h\, \delta h\, \lambda^2 = 4a^2\, 2 \sin\theta\, \cos\theta\, \delta\theta. \tag{13}$$

In our case, $\delta h = 1/Q$, and so, dividing (12) by (13)

$$Q = h \tan\theta / N\delta\theta, \tag{14}$$

where $N = h^2 + k^2 + l^2$.

This means that a diffraction line with three different indices would have three pairs of side-bands in different positions, corresponding with the substitution of

the values of h, k, and l, for h in (14). For most of the low order lines, however, the side-bands are single since the indices are either zero, when the side-band would fall on the main line, or equal. For 311 and 331, for which this is not true, we have assumed that we have measured the separation corresponding to the index 3; for 420 the difference in separation of the two sets of side-bands is less, and this is in agreement with the lack of clarity which made measurement of peak positions for this reflexion quite impossible.

That the separations of the side-bands agree with equation (14) is shown in table 1 by the fact that the values of Q calculated from them are constant within the limits of experimental error.

TABLE 1

indices	radiation	θ (degrees)	θ (radians)	Q
1 1 1	Fe	27·9	0·0037	47·8
2 0 0	Cu	25·5	0·0047$_5$	50·2
2 0 0	Fe	32·7	0·0063	51·0
2 2 0	Fe	49·8	0·0058	51·0
3 1 1	Cu	45·6	0·0060$_5$	46·1
3 1 1	Fe	63·6	0·0100	52·1
2 2 2	Fe	69·4	0·0079	56·0
3 3 1	Cu	69·8	0·0097$_5$	45·5
			Mean	50·0

(3) *Intensities of the side-bands*

The intensities of the side-bands provide less satisfactory evidence. A periodicity in scattering factor would give a constant value for the ratio of the intensity of the side-spots to the main reflexion; a periodicity in lattice parameter would make it increase rapidly with index. Measurements show that neither of these rules is obeyed. I/I_0, the ratio of the intensities of the side-bands to the main lines on a powder photograph, has an erratic increase with angle; the details are given in table 2. It might be thought that a combination of the two types of periodicity might explain the data, but the asymmetry found by Preston (1938) has not been noticed by us. Moreover, there is definite evidence against this theory.

TABLE 2

indices	I/I_0	I/NI_0	Qb/a
1 1 1	0·28	0·093	0·61
2 0 0	0·35	0·088	0·59
2 2 0	0·50	0·062	0·50
3 1 1	0·45	0·041	0·40
2 2 2	0·60	0·050	0·45
3 3 1	0·50	0·026	0·32
4 2 0	0·47	0·023$_5$	0·31

First, the photographs of the coarse-grained specimens show no satellite spots lying on the Debye lines of the main reflexions. Whatever the orientations of the

crystals, reflexions with the same value of θ should lie on these lines. This is direct evidence that the satellites corresponding to zero index have zero intensity, in accordance with the results from equation (11). Indirect evidence in support of this is given by the powder photographs, as there is no trace of higher background beneath the main lines. It is true that this would not be clearly visible, but if there were an increase of the same order as the observed satellites, it could not be missed.

Also, the intensities of the side-bands, although they do not agree with either theory, do show a reasonably smooth variation if each satellite spot is assumed to have an intensity proportional to the square of the corresponding index. If the intensities were independent of index, quite a different variation should arise. In fact, it may easily be shown that the reflexion 200, for instance, is quite incompatible with this possibility; having two zero indices, it should have four satellite spots coincident with the main line, and this is not strong enough to admit of this possibility if they were equal in intensity to the observed side-bands. This may be seen from the data given in table 2. We have also considered other sorts of possibilities, but none gives as smooth a variation as the assumption that the intensities of the satellite spots are proportional to h^2.

In this case, the intensity of the side-bands should be proportional to the sum of the squares of the indices, that is, N. Again, there is some doubt in lines of mixed indices concerning the distribution of the satellites between the main line and the side-bands; line 311, for instance, agrees better if N is taken as 9 instead of 11, suggesting that the satellites corresponding to index 1 are included in the main line.

In column 3 of table 2 values of I/I_0N are given. These are not constant as they should be if the periodic variation were entirely one of lattice parameter. Thus the values of Qb/a derived from them, and given in the next column, are also not constant; but as we have pointed out above, the variation with angle is reasonably regular.

DISCUSSION OF THE RESULTS

It is useful at this stage to summarize the conclusions that have been reached.

It has been shown definitely that (a) there is a modulation in the structure which is parallel to the cube planes, and (b) the period of the modulation is about 50 unit cells (180 A). The breadth of the side-bands would suggest that the period varies from about 30 to 70 unit cells.

The type of modulation has not be decided definitely; modulation of scattering factor fails completely to explain the experimental data, and modulation of lattice parameter, while it accounts for most of the data, does not explain completely the variation of the intensities of the side-bands with angle. Variations in scattering factor are, of course, not to be expected in an alloy of elements whose atomic numbers are so close together, but it was thought unwise to neglect completely their possibility.

It also cannot be decided whether the modulation is one, two, or three dimensional; that is, whether the maxima and minima of the modulation lie on planes

parallel to the cube faces, on lines formed by the intersection of two such sets of planes, or on points formed by the intersection of three. The former, of course, fits in well with the usual ideas of lamellar precipitation in alloys and is the easiest arrangement to envisage. It is hoped to get direct evidence by means of the electron microscope.

Since the main defect of the theory lies in the intensities it gives, and since the calculations assume that hQb/a should be small compared with unity, it is necessary to see if the defect is removed by the application of the rigorous theory. In table 2 it is seen that Qb/a is of the order of 0.5; thus hQb/a is by no means small. Nevertheless this cannot explain the discrepancy; in fact it increases it. For we have seen that the intensities of the main lines and the satellites are proportional to the squares of the different orders of the Bessel functions of hQb/a. If Qb/a is 0.5, $J_0(hQb/a)$ is nearly zero when $h = 4$, and also the number of satellites should be quite large. It is seen from figure 1 b that this is not so; the main lines are all quite strong, and the satellites show only traces of higher orders.

We are thus placed in the puzzling position that the approximate theory fits the numerical results better than the rigorous one. This is probably bound up with the lack of perfection of the modulations. We have seen that according to the breadths of the side-bands Q may vary from 30 to 70 unit cells; the lower limit may be even smaller since the intensity decreases as Q decreases. Thus although the average value of Q is 50 there may still be a large volume of the crystal for which Q is much smaller. This would contribute little to the side-bands and would leave the main lines quite strong, as they are observed to be. On the other hand, it is impossible to explain the relative intensities of the side-bands in this way since they should still increase rapidly with angle. No type of distortion that we have been able to treat mathematically will produce any change in the intensities of the side-bands; they merely change their shapes. It is possible, of course, that with increase of order part of the side-bands merges into the background, but it is not easy to account for the strength of the main lines in this way.

CONCLUSIONS

In view of the lack of certainty about the nature of the periodicity it is probably unwise to make any precise deductions from the numerical results we have given. Nevertheless, it may be pointed out that these results do fit in well with the data given by Bradley for the tetragonal structures.

The lattice parameter of the single-phase alloy is 3.5758 A; those of the two phase alloy are 3.5620 and 3.5884 A, which thus show changes of 0.013 A. In the tetragonal state the c axes differ from the single-phase spacing by about ± 0.030 A, while the a axes are unchanged; this is about the right value if the two phases are to have the same atomic volumes as in the final state. From the results given in this paper the average spacing variation in the periodic structure is about 0.032 A; this is about equal to the variation found for the tetragonal structures. Also Bradley has estimated

that the tetragonal lamellae are about 100 A thick; this is in good agreement with our results since the thickness should correspond to half the wave-length of the periodicity, that is, 90 A.

It would thus seem that the periodic structure is a first step in the process of dissociation into two phases. A complete analysis of the problem and of the changes that the alloy undergoes should throw much light on the mechanism of dissociation in this alloy and, perhaps, in alloys in general.

REFERENCES

Bradley, A. J. 1940 *Proc. Phys. Soc.* **52**, 80.
Bradley, A. J., Cox, W. F. & Goldschmidt, H. J. 1941 *J. Inst. Met.* **67**, 189.
Bradley, A. J., Lipson, H. & Petch, N. J. 1941 *J. Sci. Instrum.* **18**, 216.
Bragg, W. L. 1941 *Proc. Roy. Soc.* A, **179**, 61.
Dehlinger, U. 1927 *Z. Kristallogr.* **65**, 615.
Kochendörfer, A. 1939 *Z. Kristallogr.* **101**, 149.
Köster, W. & Dannöhl, W. 1935 *Z. Metallk.* **27**, 220.
Preston, G. D. 1938 *Proc. Roy. Soc.* A, **167**, 526.

Molecular wave functions for lithium

By C. A. Coulson, *University College, Dundee* AND
W. E. Duncanson, *University College, London*

The method of molecular orbitals has been applied to a discussion of the ground state of Li₂. Both the simple product-type functions and the determinantal functions have been used. The final energy value has an error of about 1·3 %. The changes in screening constants in passing from atomic to molecular states is discussed, and empirical rules deduced from this and other work. Provided that all the molecular orbitals used are mutually orthogonal, there is not as much electron exchange between inner and valence electrons as has often been supposed. The importance of this orthogonality in discussion of exchange is emphasized.

1. INTRODUCTION

The most fundamental and valuable step in the discussion of wave functions for atoms was the introduction by Hartree of the self-consistent field approximation. Each electron is supposed to move in a certain field due to the other electrons whose

motions we may imagine to have been 'smoothed out', or 'averaged', in the process of determining their appropriate contributions to the self-consistent potential. In a first approximation the complete wave function was merely a product of the individual atomic orbitals (Hartree approximation), but subsequently Fock showed how by taking proper account of the true determinantal symmetry, it was possible to extend the analysis so that allowance was made for electron exchange (Fock approximation). We have only to list some of the applications that have been made, using either the Hartree or Fock methods, to realize how valuable such wave functions have been—for example, electric and magnetic susceptibilities, electron velocities and shape of Compton profile, electron scattering, hyperfine structure and oscillator strengths.

The corresponding application to molecules has been developed largely through the labours of Mulliken under the title of molecular orbital approximation. But practically no numerical calculations of these wave functions have hitherto been made. The explanation is quite simple—when there are two or more centres of force we lose the central symmetry so typical of atomic structure, and the separation of variables no longer takes place. On the one hand, this introduces more variables, and on the other it requires us to evaluate some extremely formidable integrals. As a result the only calculations of this kind have been for H_2 by Coulson (1937 b, 1938), Li_2^+ by James (1935), HeH^+ by Coulson & Duncanson (1938), and methane CH_4 by Coulson (1937 a) and by Buckingham, Massey & Tibbs (1941). In this latter case the four protons were averaged over the surface of a suitable sphere so that central symmetry was artificially created; such an approximation is only practicable for relatively few molecules.

It is desirable that the method of molecular orbitals should be applied to other more complicated molecules; in this present paper we give the results of such a study with Li_2. In one sense this represents a survey of the possibilities of the molecular orbital method. We have therefore tried to make our work as useful as possible as a guide to the degree of accuracy likely to be obtained in other more complex molecules. For that reason we have used both the Hartree & Fock types of wave function so that an estimate of the loss of accuracy when using Hartree wave functions may be obtained. It became quite clear, when we tried to extend our work to N_2, that in any but the simpler molecules the Fock wave functions were quite impracticable. We have also tried to get molecular orbitals that were readily usable for other subsequent calculations. We have therefore not concerned ourselves with the detailed working of James (1934), who has also studied this molecule, even though he obtained a more accurate result for the energy. For this latter method is not pictorial since we do not assign electrons to definite orbits; also the wave functions which could not easily be extended to more complicated molecules, cannot even in this case conveniently be used for any of the subsequent calculations of other properties of the molecule such as those listed earlier. It is quite clear that in molecular problems a compromise is necessary between extreme accuracy and simplification. For that reason a qualitative understanding of the changes undergone by atomic

orbitals in the process of molecule formation is valuable. We present such a discussion in this paper.

There is also another point to consider, especially if our concern is with other properties than the energy of binding. In the calculation of such quantities as the dipole moment, or magnetic susceptibility, it is the absolute accuracy of the wave function that matters: whereas in calculating the binding energy (which is usually taken, somewhat unhappily, as the criterion of success in molecular calculations) it is the small difference in energy between the molecule and the constituent atoms that matters. This latter difference may be quite inaccurate even though the individual energies are relatively much more accurate. Thus in the case of Li_2 the total energy of the molecule relative to infinite separation of nuclei and electrons is approximately -15 atomic units (1 a.u. of energy $= 27 \cdot 08$ eV), but the binding energy is only $0 \cdot 04$ a.u. It follows that the calculated binding energy may be in error by 100 % even though the total energy is in error by only $\frac{1}{4}$ %. Even if we neglect the inner shells, which certainly contribute the greater part of the total energy, an error of 100 % in the binding energy represents an error of only 10 % in the energy of the two binding electrons.

2. Choice of wave functions

As part of the compromise between accuracy and simplicity we have used linear combinations of atomic orbitals to represent each molecular orbital. This is Mulliken's LCAO approximation, and it certainly lends itself best to further calculations. The individual atomic orbitals were of Slater type (1930). Since the binding is regarded as primarily due to the two $2s$ electrons of the Li atoms, we define the normalized atomic orbitals:

$$\psi(1s_a) = \sqrt{(c^3/\pi)}\, e^{-cr_a}, \qquad \psi(1s_b) = \sqrt{(c^3/\pi)}\, e^{-cr_b}, \left.\vphantom{\begin{matrix}1\\1\end{matrix}}\right\} \tag{1}$$
$$\psi(2s_a) = N(r_a - \alpha)\, e^{-dr_a}, \qquad \psi(2s_b) = N(r_b - \alpha)\, e^{-dr_b}.$$

r_a and r_b are the distances of an electron from the nuclei, N is a normalizing constant, α, c and d are constants to be determined later. There are reasons for supposing that c and d (especially d) may differ from the Slater values (Coulson & Duncanson 1942; Coulson 1942 a). Indeed the elucidation of this uncertainty was one of the main reasons for the present research.

According to the molecular orbital approximation the electrons in Li_2 are regarded as being: two in state $\psi(1s_a)$, two in state $\psi(1s_b)$, and two in a valence orbital $\psi(2s\sigma)$ defined by

$$\psi(2s\sigma) = N_\sigma\{\psi(2s_a) + \psi(2s_b)\}. \tag{2}$$

N_σ is a normalizing factor. If we use the Hartree approximation the full wave function Ψ_H is

$$\Psi_H = \psi(1s_a : 1)\, \overline{\psi}(1s_a : 2)\, \psi(1s_b : 3)\, \overline{\psi}(1s_b : 4)\, \psi(2s\sigma : 5)\, \overline{\psi}(2s\sigma : 6), \tag{3}$$

in which we have used a bar over a ψ to denote negative spin; its absence denotes positive spin. If we use the Fock approximation it is

$$\Psi_F = \Sigma(-1)^P P\{\psi(1s_a:1)\,\bar{\psi}(1s_a:2)\,\psi(1s_b:3)\,\bar{\psi}(1s_b:4)\,\psi(2s\sigma:5)\,\bar{\psi}(2s\sigma:6)\}, \quad (4)$$

where, as usual, P denotes any permutation of the electrons. For the sake of brevity we shall write A, B, σ for $\psi(1s_a), \psi(1s_b), \psi(2s\sigma)$ respectively. Then (3) and (4) become

$$\Psi_H = A(1)\,\bar{A}(2)\,B(3)\,\bar{B}(4)\,\sigma(5)\,\bar{\sigma}(6), \quad (5)$$

$$\Psi_F = \Sigma(-1)^P P\{A(1)\,\bar{A}(2)\,B(3)\,\bar{B}(4)\,\sigma(5)\,\bar{\sigma}(6)\}. \quad (6)$$

The energy is calculated from the usual formula $E = \int \Psi^* H \Psi \, d\tau^6 \Big/ \int \Psi^* \Psi \, d\tau^6$, where the Hamiltonian H is

$$H = \sum_{i=1}^{6} H(i) + \sum_{i>j} 1/r_{ij},$$

and

$$H(i) = -\tfrac{1}{2}\nabla_i^2 + \frac{3}{r_{ai}} + \frac{3}{r_{bi}}.$$

$H(i)$ is an operator of position of electron i only. As is usual in such calculations we are to choose values of the unknown parameters c, d, α to minimize E.

An important point arises here. If the simple Hartree calculation is to mean anything the individual orbitals must be orthogonal. For the Hartree wave function makes no allowance for the Pauli exclusion principle, and each electron will therefore try to get to the lowest state of the same symmetry, without consideration of how many electrons there may already be in that state. If, however, we compel all the orbitals to be orthogonal, this cannot occur and there will be an absence of what Van Vleck & Sherman (1935) have called 'sagging'. If we do not compel this orthogonality our calculations are meaningless. Such omissions have often been made (in slightly different form) in earlier molecular computations. For example, we make this mistake every time that we replace the inner electrons of an atom by an effective negative charge at the nucleus, and thus assume complete (or partial) screening. There is an exception to this rule of orthogonality for the atomic $1s_a$ and $1s_b$ states. The overlapping of these orbits is so small ($S_{AB} = 0.0001$ and S_{AB} always occurs with another factor of the same order of magnitude) that this condition may be relaxed for these two orbits. Similar arguments concerning orthogonality do not occur with the Fock treatment, though on grounds of simplicity in formulation and interpretation it is convenient to preserve this orthogonality among the component molecular orbitals. We shall return to this point in a later section.

In our case we have to make the $\psi(2s\sigma)$ orbital of (2) orthogonal to $\psi(1s_a)$ and $\psi(1s_b)$. This is done if

$$\alpha = \frac{24\pi(c+d)^{-4} + J_2(c,d,\rho)}{8\pi(c+d)^{-3} + J_1(d,c,\rho)},$$

where J_1 and J_2 are two integrals defined in a recent table of such integrals (Coulson 1942b), and ρ is the internuclear distance, here taken to be the experimental one, viz. 5·0 a.u. (Herzberg 1939). A word of explanation is needed for our choice of internuclear distance $\rho = 5\cdot0$ a.u. It is unlikely that our form of wave function would actually yield this for the position of minimum energy. In fact, Bartlett & Furry (1931) obtain $\rho = 4\cdot3$ a.u., and James (1934) using the full Heitler-London approximation gets $\rho = 6\cdot02$ a.u. The labour involved in computing wave functions for various values of ρ would have been prohibitive with our wave functions; and in any case we shall be concerned later with applications of these wave functions to calculations of other properties of the molecule. For this purpose it is probably better to sacrifice a small gain in energy in order to obtain a more appropriate wave function.

3. THE ENERGY

If we use the Hamiltonian of the last section with the two molecular wave functions (5) and (6) we obtain closed forms for the energies E_H and E_F. Let us write

$$H_{AA} = \int A(1)H(1)A(1)\,d\tau_1, \quad H_{AB} = \int A(1)H(1)\,B(1)\,d\tau_1,$$

$$H_{\sigma\sigma} = \int \sigma(1)H(1)\sigma(1)\,d\tau_1,$$

so that H_{AA} and $H_{\sigma\sigma}$ are the energies of the atomic $1s$ and valence $2s$ electrons under the Hamiltonian $H(1)$. We also need the r_{12} integrals. Our notation may be seen from an example. Thus let

$$[A\sigma:\sigma B] \quad \text{denote} \quad \int A(1)\sigma(2)\frac{1}{r_{12}}\sigma(1)\,B(2)\,d\tau_1\,d\tau_2.$$

Then it may be shown by straightforward analysis that

$$E_H = 4H_{AA} + 2H_{\sigma\sigma} + [\sigma\sigma:\sigma\sigma] + 4[AB:AB] + 8[A\sigma:A\sigma] + 2[AA:AA]. \tag{7}$$

Similarly

$$\begin{aligned} E_F = {}& 4(H_{AA} - S_{AB}H_{AB})(1 - S_{AB}^2)^{-1} + 2H_{\sigma\sigma} + [\sigma\sigma:\sigma\sigma] \\ & + 2\{[AB:AB] - [AB:BA] + 4[A\sigma:A\sigma] - 4S_{AB}[A\sigma:B\sigma] - 2[A\sigma:\sigma A] \\ & + 2S_{AB}[A\sigma:\sigma B]\}(1 - S_{AB}^2)^{-1} + 2\{[AA:AA] + [AB:AB] \\ & - 4S_{AB}[AA:AB] + S_{AB}^2([AA:BB] + [AB:BA])\}(1 - S_{AB}^2)^{-2}. \end{aligned} \tag{8}$$

It appears from numerical calculations with likely values of the exponents that all the terms involving S_{AB}, S_{AB}^2 and even $[AB:BA]$ may be neglected. Thus

$$E_F = 4H_{AA} + 2H_{\sigma\sigma} + [\sigma\sigma:\sigma\sigma] + 4[AB:AB] + 8[A\sigma:A\sigma]$$
$$+ 2[AA:AA] - 4[A\sigma:\sigma A] \tag{9}$$

$$= E_H - 4[A\sigma:\sigma A]. \tag{10}$$

We do not need to reproduce the details of these calculations, for they follow familiar lines, except that several quite formidable integrals were required.

4. RESULTS AND DISCUSSION

The minimum energy, using the Hartree wave function (5) and making allowance for the Coulomb repulsion between the nuclei, is given in table 1 below, together with the corresponding values for the exponents c and d. The unit of energy is the atomic unit 27·08 eV. This table also shows the corresponding values of the exponents and energy for the Li atom, calculated in the same way. The effect of the electron exchange can be found by calculating the extra term $[A\sigma : \sigma A]$ in equation (10). The appropriate values of the exponents and the energy are shown in the table.

TABLE 1. VALUES OF EXPONENTS AT MINIMUM ENERGY

	c	d	α	$E_{\text{min.}}$
Hartree wave functions:				
molecule	2·700	0·766	1·042	− 14·8072
atom	2·697	0·720	0·878	− 7·4075
Fock wave functions:				
molecule	2·70	0·81	1·008	− 14·8177
atom	2·694	0·767	0·867	− 7·4138

It is seen from this table that the advantage gained by using the Fock wave functions is approximately 0·0105 a.u. = 0·28 eV. This is, however, only a small proportion, about 1 in 1500, of the total energy. The reason why this gain is so small is that exchange between a $2s\sigma$ electron and an inner shell $1s$ electron is very small. We must expect that the difference would be larger in the case of a molecule where there were more than two electrons in the outer shells.

The small extent of electron exchange between the inner and valence shells seems to be in direct contradiction to the work of James (1934), who performs a Heitler-London calculation and lists the contributions to the energy which arise from all such possible transitions. The explanation of the divergence is soon found. For the different orbitals that James uses are not mutually orthogonal; consequently the orbitals of his valence electrons contain part of the orbitals of the inner shell. If we use determinantal wave functions for the whole molecule, this redundancy does not matter, since by subtracting rows or columns suitably we can obtain an entirely equivalent determinant in which all the functions are orthogonal. Indeed, the precise meaning to be given to James's figures for the contribution of these triple and higher exchanges is hard to define, for we could change these contributions to almost any desired extent without altering the total energy at all by simply taking new linear combinations of the same atomic orbitals. In our own case using molecular orbitals we too shall find significant contributions apparently arising from exchanges with inner and outer shell electrons unless we insist on our orbitals being orthogonal. Then equations (9) and (10) show that the interaction between inner and outer

electrons is represented by a Coulomb term $8[A\sigma : A\sigma]$ and an exchange term $-4[A\sigma : \sigma A]$. As we have shown, this latter is relatively unimportant. We conclude that this type of exchange between different shells is not very serious and may often be neglected if we take the precaution of making the various molecular orbitals orthogonal. In such a condition there are no triple and higher exchanges at all.

We may illustrate this requirement of orthogonality in a slightly different way as follows. Instead of making the molecular orbital $\psi(2s\sigma)$ defined by equation (2) as being $N_\sigma\{\psi(2s_a) + \psi(2s_b)\}$, orthogonal to the inner electrons $\psi(1s_a)$ and $\psi(1s_b)$, we could make the component atomic orbitals $\psi(2s_a)$ and $\psi(1s_a)$ orthogonal, thus using the same orbitals for the molecule as for the isolated atom (though perhaps with a different exponent in the exponential parts). Using the Hartree wave function (5) we find that the total energy drops from $-14\cdot807$ to $-14\cdot850$ a.u. The difference of $0\cdot043$ a.u., which is about 1 eV, represents the 'sagging' due to inclusion of some part of the inner shell orbital in the valence orbital, and shows how absolutely necessary it is in a Hartree treatment, and how desirable in a Fock treatment, to use orthogonal orbitals. Exactly the same situation obtains for the isolated atoms. Thus with Li, the non-orthogonal orbitals give an energy $-7\cdot7151$ a.u., whereas the orthogonal ones give only $-7\cdot4075$ a.u. The sagging is over 8 V, and the value $-7\cdot7151$ a.u. is quite meaningless.

We may make two further comments. First, that the exponent in the $1s$ wave function is not appreciably changed from its value for the separated atoms. If the exchange between these electrons and the outer ones had been as important as has sometimes been stated, we might have expected a more significant difference than we actually find. Our result, incidentally, is in agreement with the work of Buckingham et al. (1941), who found a similar conclusion for the $1s$ electrons of CH_4.

Our second comment concerns the ratio of the exponents for the valence electrons in the molecular and atomic states. It is seen from table 1 that the exponent d is larger for the molecule than for the atom, both in the Hartree and the Fock wave functions. The present writers (1942) have suggested a ratio between $1\cdot1$ and $1\cdot2$ for these exponents. A value greater than unity has been discussed by Mulliken (1932, 1940) as a condition of firm binding. A value just less than $1\cdot2$ fits the case of molecular hydrogen (Coulson 1937b). Taking the case of Li$_2$, if we use determinantal wave functions the ratio is $0\cdot81/0\cdot77$, i.e. $1\cdot06$; the same value is found for the Hartree wave functions also. Now these ratios compare the exponents in the wave function of the binding electron of the atom and the molecule when using the same type of wave function. Probably a more useful comparison would be between the molecular orbital exponent and the Slater (1930) atomic orbital exponent, since considerably more information is available concerning the latter. In the case of atomic Li the Slater exponent of the valence electron is $0\cdot65$. Using the Fock type molecular wave function as in table 1, the ratio of exponents is $0\cdot81/0\cdot65$, i.e. $1\cdot25$. This value is somewhat larger than might previously have been expected, though of the predicted order of magnitude. A reasonable assumption, therefore, to make when dealing with more complex molecules for which accurate computations are impracticable, would

appear to be to use Slater's rules for atomic wave functions, and to multiply binding electron exponents of type σ by a factor $1\cdot2$, and binding electrons of type π by $1\cdot1$, and to leave unchanged the exponents of all inner and non-bonding orbits, in passing from atomic to molecular states. This rule still leaves uncertain the case of anti-bonding orbits such as the σ^*2s of N_2. Here the available evidence from similar orbitals of H_2^+ and H_2 suggests that a small reduction in exponent may be necessary. This problem requires further elucidation.

5. Accuracy of the energy value and wave function

In the previous section we have been concerned with the relative accuracy of the Fock and Hartree types of wave function. If we are interested in the absolute accuracy of our wave functions we shall compare our values of the energy with the more laborious calculations of James (1934) and with the experimental value. Such a comparison is shown in table 2. The experimental value is found by taking twice the experimental energy of a Li atom (James & Coolidge 1936), viz. $-14\cdot9578$ a.u., and adding the observed dissociation energy D_0 which is $1\cdot14$ eV (Herzberg 1939) and correcting for the zero-point energy $0\cdot02$ eV. The percentage error with our functions is about $1\cdot3$, and even the much more complicated function of James has

TABLE 2. COMPARISON OF ENERGIES OF Li_2

experimental	$-15\cdot001$
James (1934)	$-14\cdot861$
present paper:	
Hartree-type wave function	$-14\cdot807$
Fock-type wave function	$-14\cdot818$

an error of about 1 %. Indeed, it seems hardly feasible to obtain a closer energy than this for molecules other than the very simplest, and the method that we have used in this paper does seem to represent about as good a compromise between accuracy and simplicity as any hitherto proposed. It does also show that even without any great refinements the method of molecular orbitals is able to give a good account of the wave functions even of homopolar molecules.

We may next inquire whether any improvements could easily be made in our wave function. There are two ways in which we could improve our result, but both of them destroy to some extent the pictorial representation of the electron orbits. In the first place we ought to make specific allowance for the electron repulsion terms r_{12} by introducing these co-ordinates into the wave function. In the second place, as James (1935) showed for Li_2^+, the true wave function has a greater concentration of charge between the nuclei than our present type of wave function prescribes. This could be allowed for by combining a certain amount of the atomic $2p$ orbital with the atomic $2s$ orbital in the formation of the molecular orbital of equation (2). This is equivalent to stating that the binding is not pure s binding, but a hybridized compound of s and p. Unfortunately both of these two improvements would involve considerably more calculations.

REFERENCES

Bartlett & Furry 1931 *Phys. Rev.* **38**, 1615.
Buckingham, Massey & Tibbs 1941 *Proc. Roy. Soc.* A, **178**, 119.
Coulson 1937*a* *Trans. Faraday Soc.* **33**, 388.
Coulson 1937*b* *Trans. Faraday Soc.* **33**, 1479.
Coulson 1938 *Proc. Camb. Phil. Soc.* **34**, 204.
Coulson 1942*a* *Proc. Phys. Soc.* **54**, 51.
Coulson 1942*b* *Proc. Camb. Phil. Soc.* **38**, 210.
Coulson & Duncanson 1938 *Proc. Roy. Soc.* A, **165**, 90.
Coulson & Duncanson 1942 *Proc. Camb. Phil. Soc.* **38**, 100.
Herzberg 1939 *Molecular spectra and molecular structure*, p. 489. Prentice-Hall.
James 1934 *J. Chem. Phys.* **2**, 794.
James 1935 *J. Chem. Phys.* **3**, 9.
James & Coolidge 1936 *Phys. Rev.* **49**, 688.
Mulliken 1932 *Phys. Rev.* **41**, 49.
Mulliken 1940 *J. Chem. Phys.* **8**, 234.
Slater 1930 *Phys. Rev.* **36**, 57.
Van Vleck & Sherman 1935 *Rev. Mod. Phys.* **7**, 168.

The continuous absorption of light in potassium vapour

By R. W. Ditchburn, Ph.D., *Professor of Experimental Philosophy*
Trinity College, Dublin,
J. Tunstead, *FitzGerald Scholar, Trinity College, Dublin*
and J. G. Yates, B.A.

A new type of absorption tube for the study of metal vapours is described. It is shown how the effective length of the absorbing column may be calculated. Measurements of the continuous absorption in potassium vapour, extending into the vacuum ultra-violet, are described. A separation of the atomic and molecular absorption is effected. At short wavelengths the atomic absorption appears to increase rapidly, and varies linearly with the frequency. The value of the absorption coefficient at the series limit is $1 \cdot 2\,(\pm 0 \cdot 3) \times 10^{-20}$ cm.^{2}.
    The results for the atomic absorption are in conflict with theory. It is suggested that exchange effects are operative in the absorption process and may account for the discrepancy.

1. Introduction

The study of the absorption of light in metal vapours is technically difficult because of the rapidity with which the vapours attack the windows of the aborption vessel. By using a very long tube it was found possible, in the case of caesium, to obtain a measurable amount of absorption at temperatures so low that the change in the transparency of the windows during one experiment was negligible (Braddick & Ditchburn 1934). This method cannot be applied to the other alkali metals because higher temperatures are required to produce a suitable vapour pressure.

In order to avoid the difficulty, it has been usual to place the active metal at the centre of a long steel tube whose ends are kept cool, and the metal is prevented from distilling rapidly to the ends by means of an inert filling gas. This method has the disadvantage that the concentration of the vapour at any point in the tube is not definitely known, and cannot be varied in a controllable way. Also it is necessary to use a considerable pressure of the filling gas, and it is impossible to be sure that this gas is not exerting an appreciable effect upon the absorption. The continuous absorption of potassium was measured by this method by one of us about 15 years ago (Ditchburn 1928), but owing to the difficulties we have mentioned it was not possible to make a certain and satisfactory separation of the molecular and atomic absorption. This paper describes a new technique for overcoming the difficulties, and gives the results obtained by applying it to the measurement of the absorption of potassium vapour in the region 2900–1600 A. The experiments have been interrupted owing to war conditions, but the results so far obtained appear to cover the most important points connected with the atomic absorption. They give the essential features of the variation of absorption with wave-length together with a moderately accurate determination of the absolute value of the atomic absorption coefficient at the series limit.
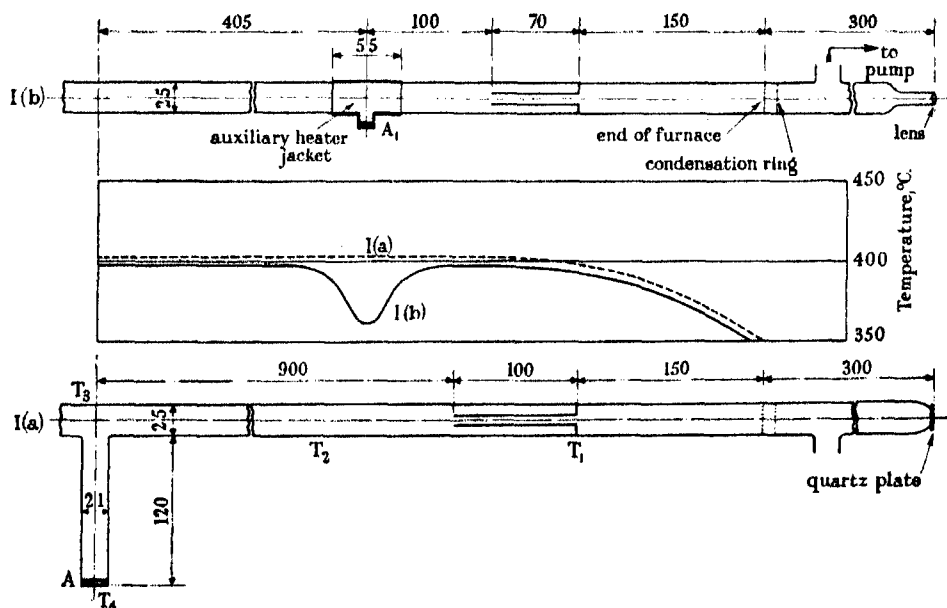
FIGURE 1. Pyrex absorption tubes and temperatures curves.
(Right-hand halves shown; dimensions in mm.; furnace omitted.)

2. THEORY OF THE METHOD

The principle of the method may be understood by reference to figure 1 which shows two different forms of absorption tube, and the temperature distribution which is maintained in each. Figure 1 (a) shows the form first used, and figure 1 (b)

that used for the final experiments. It is more difficult to arrange the heating coils and lagging to give the temperature distribution required for the second type of tube, but the accuracy of the results obtained and the simplicity of the theoretical calculations justify the extra trouble. It is intended to use this form in any future experiments. The general principle applying to both absorption tubes is that the metal is allowed to diffuse from the source A (or A 1 and A 2) to the cooler ends of the tube, but the geometry of the tubes is arranged so that the concentration of the vapour at any point can be calculated. It is found that with this type of tube a very low pressure of filling gas is sufficient to restrain the diffusion.

In order to obtain the atomic absorption coefficient it is necessary to derive an expression for the amount of light absorbed, taking account of the variation of the concentration of the vapour in different parts of the tube. We shall first make the calculation assuming that the vapour is monatomic so that the absorption coefficient is proportional to the pressure. In these circumstances the amount of light absorbed in passing through the thickness dx of the vapour at pressure p is given by

$$dI = -\beta p I dx, \tag{1}$$

where β is the absorption coefficient for unit pressure and is related to the atomic absorption coefficient σ by the expression

$$\beta = \sigma \frac{1}{760} \frac{273}{T} N, \tag{2}$$

where T is the temperature, and N is the number of molecules per c.c. at s.t.p. If the temperature is the same at all parts of the absorbing column and p is a known function of x, the expression (1) may be integrated in the form

$$\log \frac{I_0}{I} = \beta \int p \, dx. \tag{3}$$

It is convenient to calculate the equivalent length of the absorbing column, which we define to be a column of vapour at constant pressure P_m which produces the same absorption as the actual column. This length will be

$$L = \int p \, dx / P_m. \tag{4}$$

To obtain the equations required for this calculation it is desirable to consider qualitatively the process by which equilibrium is established. We start with the tube cold and the inert gas present at a certain pressure (say P). Let us assume that the system is connected to a reservoir of large volume so that subsequent changes in the temperature of the absorption tube do not alter the pressure in the reservoir. Then the total pressure at any point in the absorption tube will always be P. As the potassium is heated the partial pressure of the vapour at any point in the absorption tube will increase and that of the inert gas will decrease so as to keep their sum equal

to P. Thus the vapour drives out part of the inert gas, and when a steady temperature is reached the vapour-gas mixture is in a state of dynamic equilibrium. There is then no net movement of the inert gas in any direction, but there is a slow transfer of the metal by diffusion of the vapour through the inert gas. This type of equilibrium obtains only so long as the vapour pressure is less than P. If the vapour pressure exceeds P a bodily flow of the vapour (as opposed to a diffusion) takes place and the metal is very rapidly transferred to the ends of the tube. If a vapour pressure of the order of $0 \cdot 8P$ be used, the rate of diffusion of the vapour is still slow and the partial pressure of the inert gas in the main portion of the absorption tube is considerably less than that of the vapour. The effect on the absorption of collisions between inert-gas atoms and the vapour molecules is then negligible in comparison with that of collisions of vapour molecules amongst themselves. The usual theory for the one-dimensional problem of inter-diffusion of two gases in a tube of uniform bore leads to the relation

$$\Gamma_1 = -D\frac{dn_1}{dx}, \quad \Gamma_2 = -D\frac{dn_2}{dx}, \quad \Gamma_1 = -\Gamma_2, \tag{5}$$

where n_1, n_2 are the concentrations of the vapour and gas respectively at the point x, Γ_1 and Γ_2 are the numbers crossing unit area from left to right, at the point x, per second, and D is the diffusion coefficient. In the usual problem to which equations (5) apply both gases move along their own concentration gradients, and n_1 and n_2 are functions of the time. In our problem the vapour diffuses through the inert gas and there is no net motion of the latter. To take account of this effect we assume that at any point there is a movement of the whole system considered above, sufficient to transfer Γ_2 molecules of the inert gas from right to left. This will transfer also $\frac{n_1}{n_2}\Gamma_2$ molecules of the vapour from right to left or $-\frac{n_1}{n_2}\Gamma_1$ from left to right, so the total movement of the vapour component is

$$\Gamma^* = -\left(\frac{n_1}{n_2} + 1\right)\Gamma_1. \tag{6}$$

In a uniform tube Γ^* must be independent of x. In terms of partial pressures this becomes

$$\Gamma^* = -\frac{D}{kT}\frac{p_1 + p_2}{p_2}\frac{dp_1}{dx}, \tag{7}$$

p_1 being the pressure of the vapour, and p_2 that of the inert gas. Substituting $P - p_2$ for p_1 in (7) and integrating we obtain

$$p_1 = P - p_2 = P - p_{20}e^{\alpha x}, \tag{8}$$

where p_{20} is the value of p_2 at $x = 0$ and

$$\alpha = \frac{kT\Gamma^*}{DP}. \tag{9}$$

The form of the result is shown in figure 2. This result can be extended to the case where the tube consists of a number of different cross-sections. Neglecting end-effects, it is clear that since the total flow across any cross-section of the tube is constant, Γ^* is inversely proportional to the area of cross-section. Hence for each uniform section a solution of the form (8) is obtained with different values of α. These may be fitted together using suitable boundary conditions to secure continuity at the junctions. It is possible, however, to use the same value of α throughout if the scale of length be adjusted in each part of the tube in inverse proportion to the cross-section of that part.
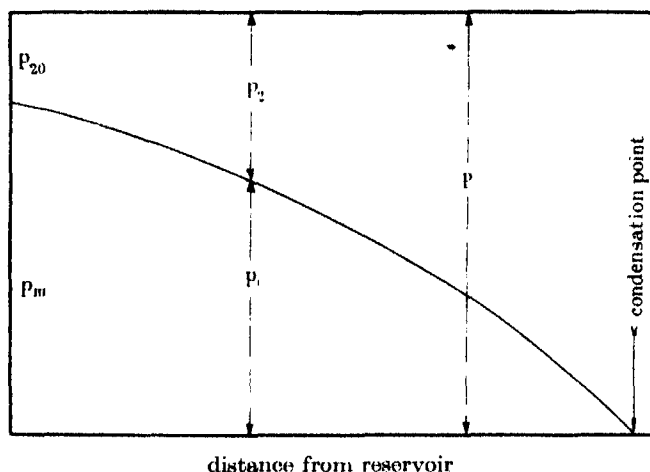
FIGURE 2. Pressure distribution in a uniform tube for the forced diffusion of a vapour through a gas.

In tube 1 *a* the flow is from the side tube to each end of the main tube. Considering each half separately, the flow can be taken as through a tube of four different cross-sections, namely, the side tube, the central main tube, the narrow section and the condensation section. It must be noted that only half the cross-section of the side tube is available for feeding one end. The value of L can now be calculated either directly from (8) using the boundary condition $p_2 = P$ at the condensation point or by a graphical method. The latter is more convenient. A suitable method is to plot the curve of the equation

$$y = P - p_{20}e^{\alpha x},$$

where y represents p_1. The areas corresponding to $\int p_1\,dx$ for different sections of the tube are obtained by a planimeter, allowance being made for the change in the scale of x. In the experiments the cross-section of the narrow portion is about one-twelfth that of the main tube. This leads to a value of α twelve times as large as that of the main tube, consequently the main fall in pressure occurs in this narrow section. With tube 1 *b* we have to deal with seven sections, the central section between

the two potassium reservoirs, each reservoir to the corresponding narrow tube, the two narrow tubes and the two condensation sections. In the practical case, nearly the whole fall of pressure is in the narrow tube and the condensation section. It is sufficient to assume that the pressure in the section between the two reservoirs is the vapour pressure corresponding to the temperature of the reservoirs, and to use the above theory to give the pressure in the remainder of the tube.

The simple theory given above involves several approximations whose effect may now be considered.

(a) *End-effect.* The one-dimensional diffusion equation does not apply in the region where an abrupt change of cross-section occurs. The correct equations are very similar to the equations for the electrical resistance at the junctions of conductors of different diameters. Applying the solution of this latter problem, it is found that the correction to the effective length of the absorption tube is negligible.

(b) *Variation of temperature along tube.* The diffusion coefficient is approximately proportional to T^2 and therefore α is approximately inversely proportional to T. The variation of temperature in the main part of the absorption tube is so small that it is sufficient to use the average value throughout. The variation in the condensation section will produce a small correction tending to increase α and hence to decrease the effective length of the tube a little.

(c) *Variation of composition along tube.* The value of D and hence of α is a function of the partial pressures, but the variation is very slow (Kennard 1938) and the effect may be neglected.

(d) *Eddy effect.* The theory would become inapplicable if large convection currents were to take place. Such convection currents are not very probable in the narrow sections of the tube. In the wide section their effect would tend to equalize the concentration of the vapour at different parts of the section. This would increase the concentration at the ends of the wide section and hence increase the rate of flow a little. Since the variation of concentration in this section given by the above theory is already small a further reduction in the variation is probably unimportant.

(e) *Molecular absorption.* In the simple theory it was assumed that the vapour is monatomic and that absorption is directly proportional to the vapour pressure. It is known that about 3 % of the vapour is in the form of diatomic molecules and that in certain regions of the spectrum these are responsible for an appreciable fraction of the whole absorption. The fraction of molecules is so small that it is not necessary to take account of effects on the diffusion equilibrium due to differences between the molecular and atomic diffusion coefficients, etc. It is, however, necessary to take account of the fact that the effective length of the tube for molecular absorption (L_M) is proportional to $\int p^2 dx$, whereas that for atomic absorption (L_A) is given by $\int p\, dx$. An investigation of this problem shows that for the higher vapour pressures used in our experiments L_M is about 5 % less than L_A and for the lower pressures

about 20 % less than L_4. The effect is thus sufficiently important to be taken into account, but it need not be accurately calculated.

3. Experiments in region 3000–2000 A

The general technique used is very similar to that previously employed (Braddick & Ditchburn 1934). Changes in technique will be described under the headings:

(a) *General arrangement.* The tube 1 a differs from the absorption tubes previously used in that the ends project from the furnace. It is thus possible to make the main support independent of the furnace. The tube is held in stands which permit lateral extension and is also supported by a single V in the middle of the furnace.

(b) *Temperature control.* The temperature is measured at points marked T in the diagram, by means of chromel-alumel couples, and the temperature of the side tube is measured by a Pt/Pt-Rh thermocouple, one junction of which is immersed in a sulphur boiling-point apparatus. The calibration is checked by determining the zinc freezing-point.

(c) *Preparation of potassium.* For most experiments potassium was prepared by redistilling ordinary potassium; for some experiments a special pure potassium obtained from Messrs Schuchardt was used. A small trace of sodium could not be removed.

(d) *Light source.* The hydrogen discharge tube previously described (Braddick 1934) is used.

(e) *Photographic.* Ilford Process plates are used for the range 3000–2200 A, and Ilford Q plates for a few experiments in which the range was extended to 1900 A. A Hilger monochromator (D 33) is used, with camera, as a spectrograph. The prism was specially selected for transparency to the region 2000–1850 A.

4. Apparatus for the vacuum ultra-violet absorption measurements

(a) *Construction.* For these experiments the absorption tube must be connected to the lamp on one side, and to the vacuum spectrograph on the other, by gas-tight joints. As the absorption tube has to be heated up through approximately 400° C, the joints must be flexible enough to take up the thermal expansion, without distortion of the optical system. The absorption tube is a complete unit being closed at the ends by waxing on 1 cm. diameter fluorite lenses.

The spectrograph is of the type described by Cario & Schmidt-Ott (1931). The end of the absorption tube fits into and is waxed to a rigid steel stand which is joined to the spectrograph through a 6 in. length of 'tombac' tubing. The spectrograph is placed on a mounting designed on geometrical principles. The latter enables the necessary adjustments to be made while preventing any relative motion between the lens on the absorption tube and the slit of the spectrograph. With this arrangement the total thermal expansion has to be taken up at the end of the absorption tube next the lamp. The latter is connected to the tube through a small length of rubber

tubing. The light from the capillary passes through this rubber tube, and is collimated by the lens on the end of the absorption tube.

(b) *Furnace and temperature measurements.* The furnace is designed so that the mean temperature has to be raised by as little as possible throughout an experiment. This reduces errors due to possible distortion of the optical system. The furnace is a single steel tube wound with nichrome ribbon over a mica insulation. The two sharp minima in the temperature curve (one shown in figure 1 b) are obtained by encircling the tube with 1 in. copper bands, from which radiator fins project. The rest of the tube has a jacket of asbestos cement whose thickness varies from point to point so as to produce the desired temperature distribution. The expansion of the steel tube controls the mean temperature automatically. This is achieved by means of the differential expansion between the furnace tube and a tube of the same length kept at room temperature by passing water through it. A system of levers magnifies the expansion and operates a relay. The temperature of the two potassium reservoirs is controlled independently of the main furnace. Auxiliary heaters are required for these points. A steel jacket is made to fit exactly over the potassium reservoir and the neighbouring portion of the absorption tube (see figure 1 b). It is wound at the ends with nichrome wire over a mica layer. The whole fits inside the main furnace. About 20 W is required to raise the temperature of the reservoirs to 10° C below the mean temperature. Temperatures are measured by six chromel-alumel thermocouples; one at each of the potassium reservoirs, one at each end—near the middle of the narrow tubes—and two at the central portion of the tube. The thermocouples on the reservoirs are pushed in between the steel jacket and the glass. The other thermocouples are wired to the outside. It has been verified previously that the thermocouples at the reservoirs read the lowest temperature in that region.

Before inserting the tube in the furnace the potassium is distilled in through a side tube which is sealed off. Helium is admitted to a pressure of 1 cm. The tube may then be warmed to run the potassium into the reservoirs. Oil-diffusion pumps with liquid-air traps are used to avoid possible contamination with mercury. The helium is admitted over charcoal cooled in liquid air.

(c) *Photometric technique.* Ilford Q plates are used. The size of plate for this spectrograph is $12 \times 1 \cdot 9$ cm. On this not more than seven spectra can be taken. Exposures are started and stopped by means of a magnetically controlled shutter. In an experiment two comparison spectra are taken with the furnace at a low temperature. Two absorption spectra are taken with the furnace at a higher temperature. On cooling two more comparison spectra are photographed. The small size and enclosed shape of the spectrograph prevents the use of grids for comparison purposes. The comparison spectra are obtained by reducing the exposure time and assuming that the reciprocity law holds. This assumption is checked in the range 2300–1900 A by a series of experiments using a Hilger E 31 spectrograph, and comparing the reductions of density due to reduction of exposure time with those due to reduction of light intensity, by means of the calibrated grids. The results of

these subsidiary experiments are applied to determine the constant p in a formula of the Schwarzschild type ($D = \log It^p$). The constant was independent of wavelength and lay between 0·95 and unity for the range of densities used. It is therefore reasonable to assume that this value of p remains unchanged in the region of the spectrum 1900–1600 A.

(d) *Experimental.* Before absorption measurements can be made the apparatus must be tested for sources of error. Such errors are:

(1) Variations in the output of the light source in the region 2000–1600 A due to 'clean-up' in the discharge.

(2) Absorption due to gases emitted from the spectrograph and the photographic plate under vacuum.

(3) Spurious 'absorption' effects due to changes in the optical alignment caused by the thermal expansion of the absorption tube.

To test for the latter error an experiment, similar to the real experiment, is done in which the temperatures are lower so that the pressure of potassium vapour is negligible. The error can be reduced by the use of a rather wide slit.

5. RESULTS

(a) *Variation of absorption coefficient with wave-length and pressure.* The absorption, $\log (I/I_0)$, as a function of wave-length is shown for several different pressures in figure 3. Curves (I), (II) and (IV) were obtained with the vacuum spectrograph; curves (V) and (VI) with the quartz spectrograph; and curve (III) is a composite curve made up of results obtained by both methods. From the curve it may be seen that in the region of overlap the agreement between the two sets of experiments is satisfactory. The shape of the curve is also in good agreement with that previously obtained by one of us for potassium in the presence of small amounts of nitrogen (Ditchburn 1928; see figure 8). The most striking feature of these curves is the very large increase of absorption towards the short wave-length end of the spectrum. The absorption shown in these curves is partly atomic and partly molecular, and before attempting any theoretical interpretation it is necessary to separate the two parts.

(b) *Separation of atomic and molecular absorption.* Under the conditions of our experiments, it may be assumed that the atomic absorption is directly proportional to the concentration and the molecular absorption to the square of the concentration. The total absorption is then given by an expression of the form

$$\rho = \sigma c + \tau c^2 \quad \text{or} \quad \rho/c = \sigma + \tau c,$$

where ρ is the measured absorption coefficient, σ is the part due to atoms, and τ that due to molecules. The concentration c is proportional to the vapour pressure divided by the mean temperature of the absorption tube. The vapour pressure is obtained, from the temperature, from the equation given by Ditchburn & Gilmour (1941). If we plot ρ/c against c we should obtain a straight-line graph, and σ and τ

can be determined from the intercept on the axis and the slope, respectively. The observations were plotted in this way, for different wave-lengths, due allowance being made for variations in the effective length of the absorbing column, and for differences in temperature. In this way the graphs of atomic and molecular absorption shown in figures 4 and 5 were obtained. The analysis shows that for most wave-lengths and pressures the greater part of the absorption is due to the atoms. It is
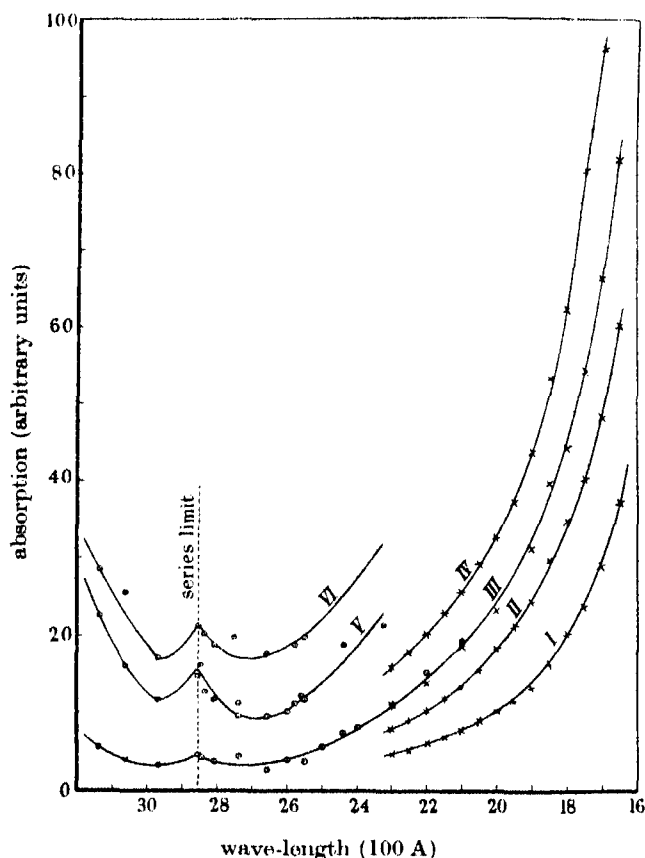
FIGURE 3. Absorption curves in the vapour-pressure range 1·6–8·5 mm.
◦ Quartz spectrograph, × vacuum spectrograph.

therefore possible to determine the atomic absorption fairly accurately, and the absorption shown in the graph is probably correct to within 20 % over most of the range. The absorption coefficient at 2300 A is more accurate than this and may be stated as $4·4 (\pm 0·5) \times 10^{-20}$ cm.^{2}; the absorption coefficient at the series limit may be given as $1·2 (\pm 0·3) \times 10^{-20}$ cm.^{2}.

The molecular absorption curve is less accurately known, since small errors in the accuracy of the results and in the method of separating will have a proportionately larger effect. It is, however, fairly certain that the molecular absorption is great in the region 3500 A and decreases to a very low value which may be zero in the

region 2500–2000 A, and increases again in the far ultra-violet region. Accepting the usual estimates for the proportion of molecules present it appears that the molecular absorption in the region 3100 A is very high, being of the order 3×10^{-18} cm.².

(c) *Shape of absorption curve near series limit and pressure effect.* In the experiments on caesium it was found that the absorption appeared to increase very sharply on the long-wave side of the series limit (Braddick & Ditchburn 1934). Close examina-
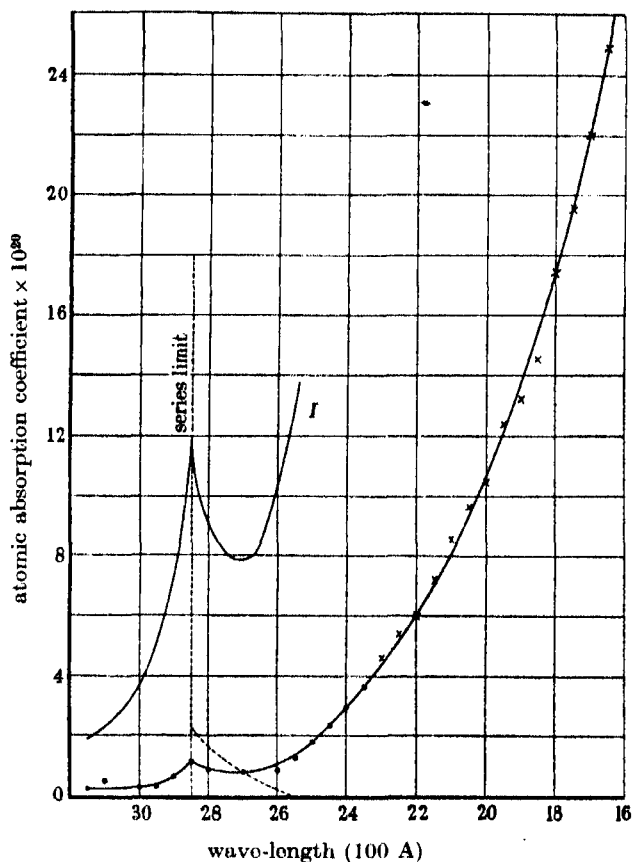
FIGURE 4. Atomic absorption curves. Dotted line theoretical.
In I, scale of *y* multiplied by 10.

tion of the photometer traces suggests that this effect probably exists in potassium, but is much weaker than in caesium. The shape of the curve near to the potassium series limit is very similar to that found for caesium in the presence of a considerable pressure of gas, and is possibly due to some interaction between the potassium atoms. The presence of the sodium absorption line 2853·9 A prevented us from determining whether the position of the maximum absorption coincides exactly with the theoretical series limit 2856·6 A.

Nearly all the experiments were done with helium as the filling gas and the pressure of helium in the tube was of the order 5 mm. According to results previously obtained on caesium (Ditchburn & Harding 1936) this amount of filling gas should exert quite negligible effects on the absorption. A very few experiments done with neon and argon as filling gas at pressures of the order of 3 cm. indicated that these did not produce any large effect; although there was probably a small reduction in the amount of absorption.
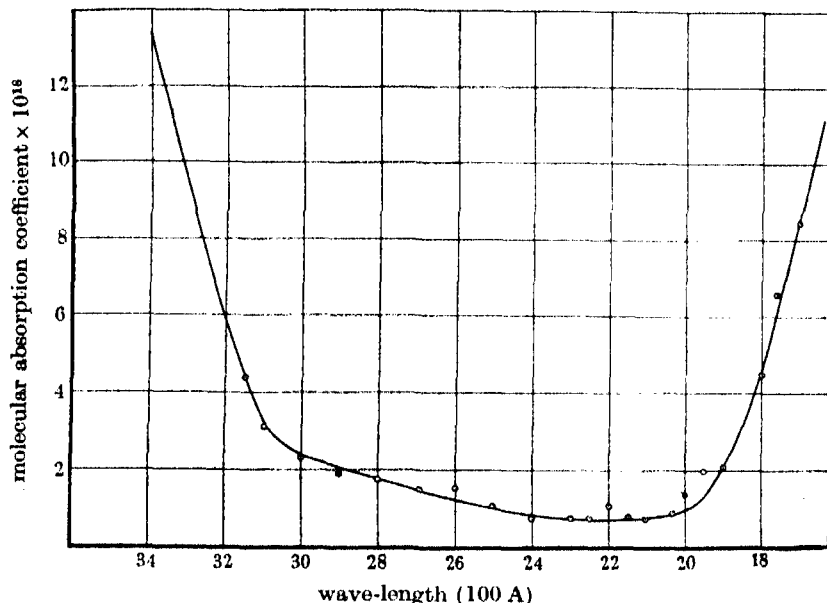
FIGURE 5. Molecular absorption curve.

6. DISCUSSION

The atomic absorption curve is in agreement with the curve obtained by the direct method (Lawrence 1925) as far as the wave-length 2300 A. The later photoelectric measurements of Lawrence & Edlefsen (1929), using the space-charge ionization spectrum, indicated the existence of a definite maximum in the curve in the region 2300 A; whereas our curves do not indicate such a maximum. Other evidence has shown that the theory of the space-charge detector is more complicated than was originally thought, and a very small variation in the efficiency of the space-charge detector would be sufficient to explain the discrepancy (McFadden 1942). It appears fairly certain that our curves give the absorption correctly and that this absorption is all due to the photoelectric process. The theoretical calculation of Phillips (1932) led to the result that the absorption should be $2 \cdot 2 \times 10^{-20}$ cm.^{2} at the series limit and should decrease monotonically, as shown in the dotted line in figure 4. It is obvious that the experimental results are in direct conflict with the theoretical calculations on this point. Early in 1939 calculations of the continuous spectrum of potassium were being carried out under the direction of Professor D. R. Hartree using the

differential analyser and taking into account exchange forces. It was hoped that these might have cleared up the discrepancy, but unfortunately it was not possible to complete the work. Recent calculations of continuous absorption coefficients (Bates 1939) have shown that for atoms like oxygen, fluorine and neon the absorption does not decrease monotonically from the series limit. Calculations on exchange effects in Ca and Ca⁺ (Bates & Massey 1941) show that exchange effects materially alter the shape of the absorption curve.
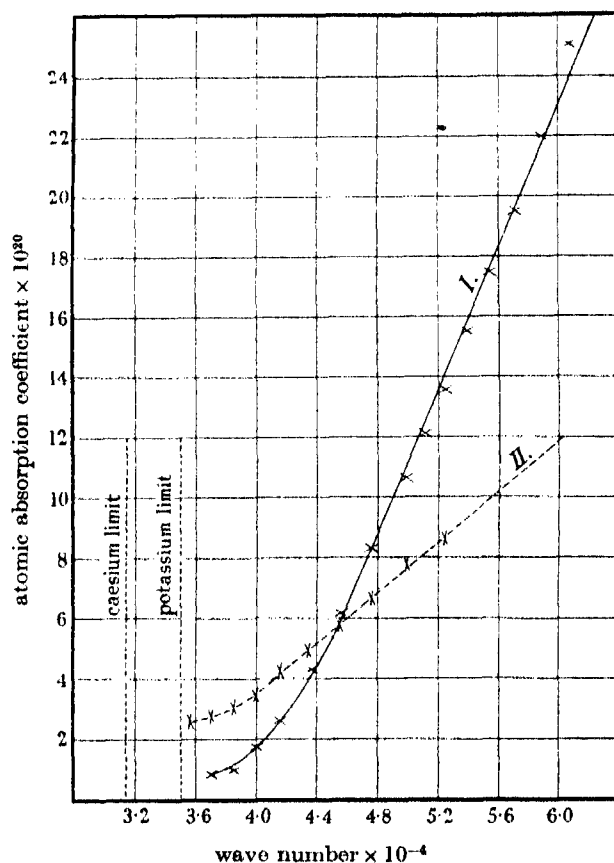
FIGURE 6. Absorption coefficient plotted against wave number. I Potassium. II Caesium. (Ordinate-scale applies to potassium. Ordinates for caesium have been reduced by a factor of three.)

In the case of caesium it was similarly found experimentally that the absorption did not decrease monotonically from the series limit, but over the portion of the spectrum measured the increase was not so rapid as in the case of potassium. For caesium it was possible that the total integral to the absorption curve (which is proportional to the f value for the continuum) could be equal to that calculated theoretically; though this would only be so if the absorption decreased very rapidly just beyond the wave-length at which the observations finished. The present results

for potassium indicate that the f value obtained for the continuum from the experimental results is many times greater than that predicted theoretically. This suggests very strongly that some process involving an electron other than the valence electron is operative. In this connexion it is probably significant that the absorption measurements for sodium and lithium (which have no incomplete inner shells) are in agreement with the theoretical calculations. For caesium and potassium in the region of short wave-lengths the absorption coefficient appears to be proportional to the frequency. In figure 6 it may be seen to what extent this empirical law holds.

REFERENCES

Bates, D. R. 1939 *Mon. Not. R. Astr. Soc.* **100**, 25.
Bates, D. R. & Massey, H. S. W. 1941 *Proc. Roy. Soc.* A, **177**, 281.
Braddick, H. J. J. 1934 *Proc. Camb. Phil. Soc.* **30**, 355.
Braddick, H. J. J. & Ditchburn, R. W. 1934 *Proc. Roy. Soc.* A, **143**, 472.
Cario, G. & Schmidt-Ott, H. D. 1931 *Z. Phys.* **69**, 719.
Ditchburn, R. W. 1928 *Proc. Roy. Soc.* A, **117**, 486.
Ditchburn, R. W. & Gilmour, J. C. 1941 *Rev. Mod. Phys.* **13**, 310.
Ditchburn, R. W. & Harding, J. 1936 *Proc. Roy. Soc.* A, **157**, 66.
Kennard, E. H. 1938 *Kinetic theory of gases*, 1st ed. p. 198. McGraw-Hill Book Co.
Lawrence, E. O. 1925 *Phil. Mag.* **50**, 345.
Lawrence, E. O. & Edlefson, N. E. 1929 *Phys. Rev.* **34**, 1056.
McFadden, T. 1942 Private communication, Queen's University, Belfast.
Phillips, M. 1932 *Phys. Rev.* **39**, 905.

Penetrating non-ionizing cosmic rays

By L. Jánossy and G. D. Rochester

An account is given of experiments proving the existence of a penetrating non-ionizing component of cosmic radiation at sea-level. It is shown that the radiation has a mean range of approximately 10 cm. in lead, and is probably the same as the penetrating non-ionizing radiation discovered by Rossi and Regener in experiments at 4300 m. above sea-level. An argument is brought forward indicating that the radiation might consist of neutrons.

I. Introduction

The main part of the non-ionizing component of cosmic radiation is known to consist of photons. The photon component was investigated by an anticoincidence method by Jánossy & Rossi (1940) and Regener (1940). The present paper deals with experiments which are an extension of the photon experiment. They give

evidence for the existence at sea-level of a non-ionizing component far more pene-trating than photons. The intensity of this non-ionizing component at sea-level is so small that it could not have been detected by Jánossy & Rossi.

A short account of some of our results was given elsewhere (Jánossy & Rochester 1941). Experiments with similar results were carried out at 4300 m. above sea-level by Rossi & Regener (1940).

II. THE EXPERIMENTAL METHOD

Consider the counter arrangement shown in figure 1. An anticoincidence (BC–A) is defined as a coincidence of the counters B and C not accompanied by a discharge of any of the anticoincidence counters A. An anticoincidence can be due to one of the following processes: (1) a non-ionizing ray produces in s an ionizing secondary which discharges the counters B and C. As this is the only kind of anticoincidence we wish to observe, such an anticoincidence is called a genuine anticoincidence.
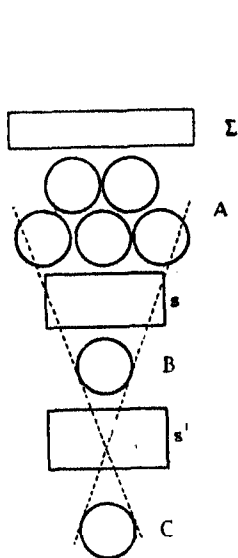
FIGURE 1. Scheme of anticoin-cidence arrangement.
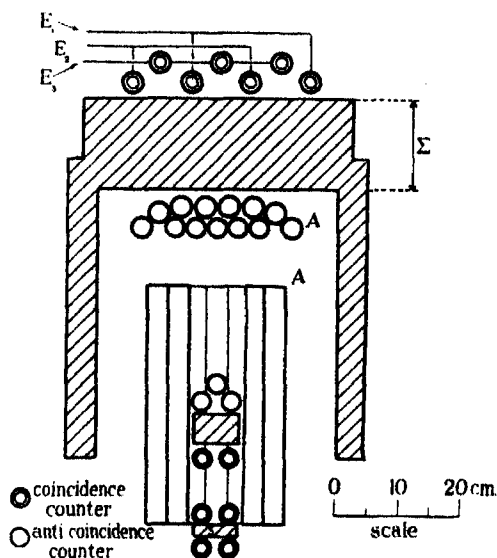
FIGURE 2. The experimental outlay, cross-section.

(2) An ionizing particle enters s without crossing any of the counters A and is scat-tered in such a direction as to produce a coincidence BC, or it produces a secondary in s which discharges the counters B and C. (3) A shower from the side which dis-charges B and C but not A. (4) An ionizing particle which passes through A, B and C but happens not to discharge A due to the inefficiency of the counters. (5) A casual coincidence between the counters B and C which is not accompanied by a discharge of A. (6) A particle which comes from below and is absorbed in s. Types of anti-coincidences 2–6 will be called spurious anticoincidences.

The experiment was carried out with the counter arrangement reproduced schematically in figures 2 and 3. The arrangement consists of a threefold coincidence set of counters (*BCD*), and an anticoincidence set of counters (*A*), containing 76 counters in parallel. The anticoincidence counters shield the coincidence system from all directions except from below. An account of the construction and the efficiency of the counters will be published elsewhere (Rochester & Jánossy 1943).
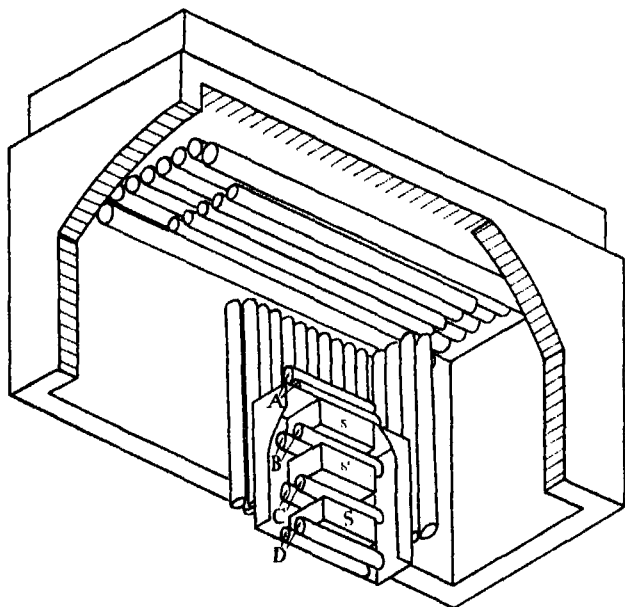
FIGURE 3. The experimental outlay, seen from below.

The threefold set *BCD* was surrounded by 5 cm. of lead on five sides to cut out photons and that this was accomplished may be shown as follows. The rate of anti-coincidences observed with the top part of the absorber removed (i.e. $\Sigma = 0$) was 2·4 anticoincidences per hour and therefore the rate due to photons alone must have been smaller. The probability of a photon penetrating 5 cm. of lead without encounter is less than $\frac{1}{2}$ %, hence the rate of anticoincidences due to photons must have been 0·01 c. per hr. or less. This rate is negligible compared with those observed (see table 2).

Since the rate of penetrating non-ionizing agents proved to be very small, great care had to be taken to reduce the rate of spurious anticoincidences to a minimum. Eventually the rate of spurious anticoincidences was reduced to less than 0·04 % of the rate of threefold coincidences.

This high efficiency was obtained in two ways:

(1) By reducing the duration of the anticoincidence counter pulses to prevent the overlapping of pulses. This was obtained by using small counter coupling units: a counter leak of 20,000 Ω and a coupling condenser of 100 P.F. In order not to miss

discharges giving rise to very small pulses the input pulses were amplified by a three-stage, resistance-capacity amplifier. The amplified pulses and the coincidence pulses were mixed by an anticoincidence amplifier of the type described by Rossi and his coworkers (1940). The efficiency of the anticoincidence amplifier was 100 %.

(2) By arranging the anticoincidence counters in such a way that a particle reaching the coincidence counters or the absorber s had, in general, to pass through two or more anticoincidence counters, though small regions could only be covered singly because of the finite thickness of the counter walls (~ 0.5 mm.). The rate of spurious anticoincidences due to this type of leakage is estimated as follows:

Lack of efficiency of a single counter (observed) ... ... $0.7 \% = 100a$

Fraction of the solid angle subtended by BCD covered
only singly by anticoincidence counters (observed) ... $2.3 \% = 100b$

Total leakage ... ... $100(a^2 + ab) = 0.016 \%$
$= 0.11$ (c. per hr.)

Thus the estimated rate of direct leakage is of the order of the smallest rates observed (see tables 1 and 2).

The rate of anticoincidence pulses was of the order of 700 c. per sec. The anticoincidence amplifier was designed to block the recording of coincidence counts for about 200 μsec. after each anticoincidence pulse. Thus the coincidence amplifier was blocked for about 0.14 sec. per sec. due to discharges of the anticoincidence counters. This blocking resulted in a loss of about 14 % of the anticoincidences. The numerical value of this blocking ratio was determined by applying artificial pulses to the coincidence input of the amplifier and by recording the fraction of these pulses not getting through the amplifier.

Tests carried out daily showed that the blocking ratio remained remarkably constant.

The space below the coincidence counters was not covered by anticoincidence counters because secondaries produced in s by non-ionizing radiation would have discharged the anticoincidence system *after* having passed through the coincidence counters and no genuine anticoincidences would have been recorded.

Since the bottom was not covered, spurious anticoincidences due to ionizing particles moving upwards and being stopped in s might be expected. We note that the rate of these spurious anticoincidences would be strongly reduced when s is shifted from its original position to the lower position s' (figure 1). An observed decrease in the anticoincidence rate when s was moved to s' would therefore not necessarily have been an indication of the presence of non-ionizing radiation. For this reason we preferred not to change the position of any of the absorbers inside the anticoincidence system throughout the main part of the experiment. We took, however, a series of readings with the lead s in the two positions s and s' and the results are given in table 1. The change of the anticoincidence rate shown in table 1

is no evidence for the existence of non-ionizing radiation but the lower rate can be regarded as an upper limit for the rate of direct leakage through A.

TABLE 1. ANTICOINCIDENCES AS A FUNCTION OF THE POSITION OF THE LEAD s

$\Sigma = 5$ cm. Pb; $s = s' = 3\cdot5$ cm. Pb

position	anticoincidences $(BCD–A)$ (corrected for random coincidences BCD, A) rate (c./hr.)
s	$0\cdot61 \pm 0\cdot07$
s'	$0\cdot26 \pm 0\cdot04$

III. THE EXPERIMENTAL RESULTS AND THEIR INTERPRETATION

(1) The rate of anticoincidences observed with $s = 5$ cm. Pb and $\Sigma = 5$ cm. Pb was $0\cdot533 \pm 0\cdot039$ anticoincidences per hour. The most direct way to show that a large fraction of these anticoincidences was due to non-ionizing radiation appeared to be to absorb the radiation by placing absorbers on top of the arrangement. By placing absorbers in the position Σ a strong reduction of the rate of anticoincidences was in fact observed (see table 2, and figure 4).

TABLE 2. ANTICOINCIDENCES AND COINCIDENCES AS A FUNCTION
OF THE THICKNESS OF THE ABSORBER Σ

Σ lead cm.	$(BCD–A)$ total count	anticoincidences $(BCD–A)$ (corrected for random coincidences BCD, A)		coincidences (BCD)	
		time hr.	rate c./hr.	time hr. min.	rate c./hr.
0·0	27	16·3	2·4 $\pm$ 0·4	—	—
5·0	193	469·2	0·533 $\pm$ 0·039	4.18	711·6 $\pm$ 1·3
12·5	155	445·8	0·451 $\pm$ 0·036	4.17	661·8 $\pm$ 1·3
25·0	89	407·7	0·283 $\pm$ 0·030	5.22	616·2 $\pm$ 1·1
5·0 + 30·0 cm. Al	119	427·3	0·362 $\pm$ 0·034	5.38	643·2 $\pm$ 1·1

Though the absolute value of the effect given in table 2 is very small, we are confident that it is real for two reasons: (1) the relative effect is large, being of the order of 100 %; (2) the results given in table 2 represent the average values of about 100 individual readings spread out over a period of three months. The internal consistency of the individual readings is very good, as can be seen from table 3 where the individual readings for $\Sigma = 5$ cm. Pb are collected. The number of each reading is given in column (1), from which it is seen that the eighteen readings for $\Sigma = 5$ cm. Pb are uniformly distributed among the seventy-two readings taken for all thicknesses

of the absorber Σ. In column (5) are given the values of $(\Delta N)^2$, the squares of the deviations from the mean value. Since the expectation value of $(\Delta N)^2$ is N, $\Sigma(\Delta N)^2$ is expected to be of the order of ΣN. It is seen from table 3 that $\Sigma(\Delta N)^2 = 125\cdot2$ while $\Sigma N = 193$, showing that the fluctuations are normal. The fluctuations for the other sets of readings for different values of Σ are also normal.
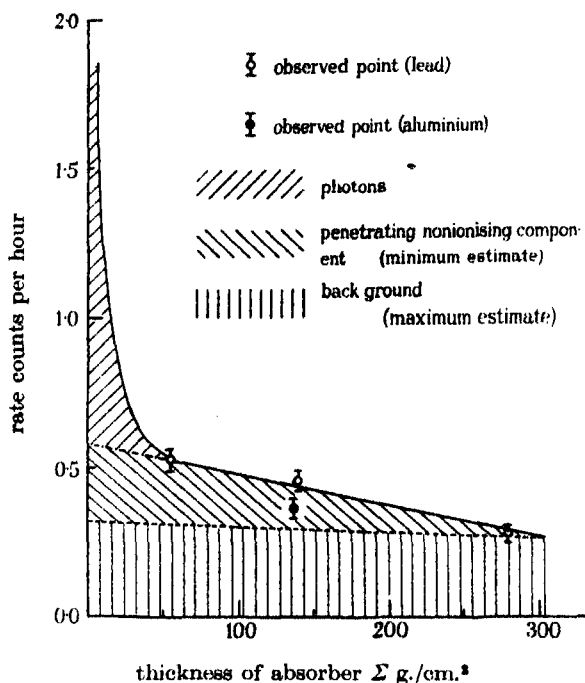
FIGURE 4. Graph of the experimental results.

(2) We interpret the difference of the rates of anticoincidences observed with $\Sigma = 5$ cm. and $\Sigma = 25$ cm. Pb as follows: for $\Sigma = 5$ cm. some non-ionizing rays penetrate the absorber Σ without giving rise to ionizing secondaries and thus reach the absorber s without setting off A. Some of these rays produce ionizing secondaries in s which give rise to genuine anticoincidences ($BCD-A$). For $\Sigma = 25$ cm., however, most of the non-ionizing rays do not give rise to anticoincidences because they are either absorbed or give rise to ionizing secondaries which discharge A. Thus we assume that the excess of the rate of anticoincidences with $\Sigma = 5$ cm. Pb over the rate with $\Sigma = 25$ cm. Pb is due mainly to secondaries to non-ionizing agents with ranges between 5 and 25 cm. Pb. To justify this interpretation we have to show that the spurious anticoincidences cannot be much affected by the variation of the thickness of the absorber Σ. We will discuss the various processes giving rise to spurious anticoincidences separately.

(i) *Direct leakage.* The rate of spurious anticoincidences due to leakage is a

constant fraction of the total rate of threefold coincidences. This ratio must be equal to or smaller than

$$r = \frac{\text{rate of anticoincidences } \Sigma = 25 \text{ cm.}}{\text{rate of threefold coincidences } \Sigma = 25} = \frac{0 \cdot 28}{616} = 0 \cdot 0005$$

(see table 2). Since the decrease of threefold coincidences, BCD, while Σ is increased from 5 to 25 cm. Pb is 96 c. per hr. (see table 2, last column), the corresponding change of the rate of anticoincidences by leakage must be $96 \times 0 \cdot 0005 = 0 \cdot 048$ c. per hr., which is small compared with the total effect. The actual decrease due to leakage is probably even smaller, since it is unlikely that all anticoincidences with $\Sigma = 25$ cm. Pb are due to leakage.

TABLE 3. INDIVIDUAL READINGS FOR $\Sigma = 5$ CM. PB AND
$s = 5$ CM. PB AND THEIR FLUCTUATIONS

reading (1)	time hr. min. (2)	count N (3)	deviation from mean ΔN (4)	square of deviation $(\Delta N)^2$ (5)
1	74.50	25	$-5 \cdot 7$	32·2
4	23.00	10	$+0 \cdot 5$	0·3
7	22.20	8	$-1 \cdot 2$	1·4
10	22.00	14	$+5 \cdot 0$	25·0
12	23.00	8	$-1 \cdot 4$	2·0
14	18.45	7	$-0 \cdot 7$	0·5
16	21.30	4	$-4 \cdot 8$	23·0
20	23.50	13	$+3 \cdot 2$	10·2
24	23.05	9	$-0 \cdot 5$	0·3
28	23.30	9	$-0 \cdot 6$	0·4
39	23.10	9	$-0 \cdot 5$	0·3
47	21.35	8	$-0 \cdot 9$	0·8
51	19.25	8	$0 \cdot 0$	0·0
58	22.05	10	$+0 \cdot 9$	0·8
62	22.00	13	$+4 \cdot 0$	16·0
66	23.53	12	$+2 \cdot 2$	4·8
69	44.45	17	$-1 \cdot 4$	2·0
70	16.25	9	$+2 \cdot 3$	5·3
totals	469.08	193		125·3

$\Sigma(\Delta N)^2 = 125 \cdot 3$ (calculated from 18 independent readings).
Average $\Sigma(\Delta N)^2 = \Sigma N = 193$ (theory).

(ii) *Particles travelling upwards which are stopped in s.* The rate of these anti-coincidences should not be affected at all by the thickness of the absorber Σ.

(iii) *Casual coincidences.* Only a small fraction of the casual threefold coincidences are recorded as anticoincidences. The largest number of casual anticoincidences are due to the overlapping of an anticoincidence pulse $CD-A$ with an anti-pulse $B-A$. The rate of these is estimated to be $0 \cdot 006$ c. per hr. which is negligible. Therefore any change of the rate due to a change of Σ must also be negligible.

Thus we conclude that the change of the rate of anticoincidences with Σ cannot be accounted for except by assuming a non-ionizing radiation.

(3) The difference between the rates of the anticoincidences with $\Sigma = 12 \cdot 5$ cm. Pb and $\Sigma = 25$ cm. Pb is $0 \cdot 17 \pm 0 \cdot 05$ c. per hr. This difference exceeds more than three times the standard error and can be regarded as real. We conclude therefore that an appreciable fraction of the non-ionizing radiation penetrating 5 cm. Pb also penetrates $12 \cdot 5$ cm. Pb. Thus the radiation observed is much more penetrating than photons.

It is difficult, however, to make a reliable estimate of the average range of the radiation since it is not clear if the anticoincidences observed with $\Sigma = 25$ cm. Pb are all spurious. (Due to a breakdown we were unfortunately prevented from extending the absorption curve up to 50 cm. Pb.) In any case it appears reasonable to assume a range of the order of 10 cm. Pb.

To obtain information on the dependence of the absorption process on the atomic number of the absorber we carried out observations with $\Sigma = 5$ cm. Pb + 30 cm. Al. 30 cm. Al is about equivalent to $7 \cdot 5$ cm. Pb. The results of this experiment are also given in table 2. The rate of anticoincidences with the aluminium absorber is distinctly lower than the rate with $\Sigma = 5$ cm. Pb only. Thus the aluminium gives rise to some absorption. It is, however, not clear if the difference between the rate for aluminium and the rate for 25 cm. Pb is significant or not. All we can conclude is that the observations are not incompatible with the assumption that the absorption is mass proportional although there is a slight indication that aluminium absorbs more strongly than lead.

(4) A further experiment was made to find out if the non-ionizing radiation is accompanied by showers falling on the absorber Σ. A sensitive shower detecting arrangement was placed above the absorber Σ and readings of $(BCDE_1E_2E_3-A)$ were taken. No coincidence of this type was recorded in 473 hr. (see table 4). Thus the non-ionizing particles are not accompanied by showers.

TABLE 4.　READINGS OF $(BCDE_1E_2E_3-A)$

Σ cm. Pb	time hr. min.	count
5·0	86.15	0
12·5	167.13	0
25·0	82.25	0
5·0 + 30·0 cm. Al	137.00	0
total	472.53	0

(5) The decrease of the rate of anticoincidences as given in table 2 is $0 \cdot 25$ anticoincidences per hour, $0 \cdot 035 \%$ of the coincidences rate BCD. This figure does not represent, however, the actual ratio between the intensities of ionizing and penetrating non-ionizing components for two reasons. In the first place, the secondaries of a non-ionizing ray must penetrate between $2 \cdot 2$ and $7 \cdot 2$ cm. Pb in order to produce

a threefold coincidence *BCD* and to be recorded. Assuming the secondaries are mesons, a minimum energy of approximately 10^8 eV will be required, and only non-ionizing agents exceeding the minimum energy will be recorded. In the second place, a non-ionizing particle can only be recorded if it is not absorbed in 5 cm. Pb above the anticoincidence counters *A* but is absorbed in the 5 cm. thick absorber *s*. Writing μ for the absorption coefficient and x for the thickness of Σ or *s*, the probability of one particle being absorbed in *s* after passing through Σ is

$$P = e^{-\mu x}(1 - e^{-\mu x}).$$

The value of P is always small as can be seen from table 5. Thus it is safe to assume that not more than one out of four non-ionizing rays gives rise to an anticoincidence. The penetrating non-ionizing component is thus expected to exceed $4 \times 0.035 \% = 0.14 \%$ of the total cosmic-ray intensity near sea-level. It seems reasonable to assume that the non-ionizing component does not greatly exceed this value, and that therefore 0.14% represents the correct order of magnitude of the intensity of the non-ionizing component.

TABLE 5

$1/\mu$ (cm. Pb)	$P \%$ ($x = 5$ cm. Pb)
5	23.2
10	23.9
15	20.3
20	17.3

IV. DISCUSSION

(1) An experiment similar to the one described here was carried out by Rossi & Regener on Mount Evans 4300 m. above sea-level. They reported a decrease in the rate of anticoincidence of $0.31 \pm 0.09 \%$ of the vertical intensity, while Σ was increased from 2.5 to 10 cm. Pb.

It seems reasonable to compare the Mount Evans value with half of the decrease of the anticoincidence rate we obtained by increasing Σ from 5 to 25 cm. Pb. Assuming the ratio of the absolute cosmic-ray intensities between sea-level and Mount Evans to be 1 : 2.3, the ratio of the intensities of the non-ionizing radiations is estimated as

$$\frac{0.5 \times 0.035}{2.3 \times 0.31} \approx \frac{1}{40}.$$

This value represents a very rough estimate only, especially as our arrangement of the absorbers differs considerably from that of Rossi & Regener.

The mass-equivalent of the atmosphere between Mount Evans and sea-level is $H = 400$ g. per cm.^{2}. The absorption coefficient of the radiation can be expressed as $\mu = (\log_e 40)/H$ and the average range (R) is

$$R = 1/\mu \approx 100 \text{ g. per cm.}^2.$$

This value is in as good agreement as one could expect with the range observed at sea-level for lead. The agreement between these two values favours the assumption that the absorption of the non-ionizing radiation is mass proportional.

(2) The question of the nature of the radiation arises. The most obvious suggestion is that it consists of neutrons or of neutrettos; but as little is known of the behaviour of energetic neutral particles from the theoretical point of view, little can be stated definitely.

In order to give a possible interpretation of the observations we consider the following processes. A fast neutron suffering a head-on collision with a proton inside a nucleus can transfer its whole momentum to the proton. According to Heisenberg (1937) the mean free path of a neutron which is stopped according to such a process is 12·5 cm. Pb. This process alone does not account for our observations, since the reverse process of protons producing neutrons should make the neutrons reappear. It has been suggested by Jánossy (1942), however, from observations on penetrating showers, that fast protons traversing matter are quickly stopped by interaction with nuclear Coulomb fields, giving rise to penetrating showers. These two processes together would give for the mean free path of the neutral radiation a value of about 10 cm. lead in agreement with the present observations.

The above hypothesis was put forward by us a short time ago (Jánossy & Rochester 1941). Since then we have carried out experiments supporting the view that the non-ionizing penetrating component is at least partly responsible for the production of penetrating showers. These experiments have been reported briefly (Jánossy & Rochester 1942).

REFERENCES

Heisenberg, W. 1937 *Abh. sächs. Akad. Wiss.* **89**, 369; *Naturwissenschaften*, **25**, 749.
Jánossy, L. 1942 *Proc. Roy. Soc.* A, **179**, 361.
Jánossy, L. & Rochester, G. D. 1941 *Nature, Lond.*, **148**, 531.
Jánossy, L. & Rochester, G. D. 1942 *Nature, Lond.*, **150**, 633.
Jánossy, L. & Rossi, B. 1940 *Proc. Roy. Soc.* A, **175**, 88.
Regener, V. H. 1940 *Ric. Sci.* **11**, 66.
Rochester, G. D. & Jánossy, L. 1943 *Phys. Rev.* **63**, 52.
Rossi, B., Jánossy, L., Rochester, G. D. & Bound, M. 1940 *Phys. Rev.* **58**, 761.
Rossi, B. & Regener, V. H. 1940 *Phys. Rev.* **58**, 837.

Some remarks on the statistics of binary systems

By Gregory H. Wannier

State University of Iowa

This paper discusses the results obtained by K. Fuchs concerning a binary body-centred cubic system. It points out that the method used should and does partly cover both the two-phase system and the case of superstructure. It cautions also against acceptance of the quantitative conclusions obtained. When tested with the help of exact results available the method is found to contain important errors both of type and magnitude.

In an article on the same subject, K. Fuchs (1942, quoted in the following as F) has applied the cluster method of statistics to the problem of two intermingling elements sharing the same lattice. The calculation indicates that the cluster method has a wider range of application than other rigorous procedures, but just for this reason it is important to dispel some erroneous ideas which the paper tends to give to the reader.

The author remarks that 'the critical temperature corresponding to the appearance of superstructure is not given by these equations'. This statement is not correct for the temperature T_c as calculated in (7·5). This temperature is the critical temperature for 50 % mixture; one can verify, by a simple statistical argument, that if the lattice can be divided into α- and β-sites with no two α- (or β-) sites as direct neighbours, a 50-50 mixture will have the same thermal properties if the interaction V is replaced by an interaction of equal magnitude, but opposite sign. This applies to the body-centred lattice treated in F. It is possible to trace this symmetry partially through the calculations. It means that for the concentration $c_a = \frac{1}{2}$, the free energy A, as defined through (3·14), must be even in V, provided the energy zero is fixed so that

$$V_{aa} + V_{bb} + 2V_{ab} = 0.$$

A short calculation shows that this means

$$G_0(\tfrac{1}{2}, \beta') - G_{01}(\tfrac{1}{2}, \beta')$$

is even in V and hence in $\xi = V/kT$. This can be verified directly for the lower powers of ξ if the form (A. 7) is used rather than (A. 5) and the expressions (A. 10) are substituted into it. Below the critical temperature this tracing seems to be somewhat more difficult to carry through.

It follows that the asymmetry apparent in (A. 14) has been accidentally introduced by the equations picked for the determination of T_c. These equations are (4·3) and (7·1) both of which have (A. 13) as a consequence. Equation (4·3) may break down in the case of superstructure either because there is no phase separation at all and T_c is an isolated singular point, or else because there are more than two phases present

involving perhaps the 25 % superstructure; in this case the identification of (3·15) and (3·17) is no longer possible. Equation (7·1), on the other hand, is only one of two possibilities as was pointed out by Born and Fuchs (1938, p. 408). The second possibility, namely that $G(c_a, \beta)$ is singular, is unfortunately not as convenient for numerical calculations.

The positive evidence of the paper as to the nature and location of the temperature singularities is also open to serious doubt until further corroborative evidence can be obtained. The method used is rigorous in principle, but before numerical results can be obtained the functions involved have to be replaced by their power series. Fortunately there are now some rigorous results available with which the results in F can be brought in comparison.

These results were obtained for the so-called Ising model of ferromagnetism (Kramers and Wannier 1941; quoted in the following as KW). The binary system is identical with it for all compositions below the temperature singularity and for the 50 % composition above it. In fact figure 1 in F, if cut in two and turned by 90°, just gives the spontaneous magnetization against temperature in the Ising model. No exact results are available yet for a three-dimensional case as the one treated in F, but they exist for a number of two-dimensional problems (Onsager, private communication), particularly the square net. With the notation used in F we get from KW, eq. (31)

$$\sinh \tfrac{1}{2}\xi = \sinh (V/2kT_c) = \pm 1.$$

In addition closed form expressions of the free energy and the specific heat as functions of temperature are now available (Onsager, private communication). The specific heat has as its dominant feature that it is infinite at T_c, tending to infinity in a rather symmetric manner both from higher and from lower temperatures. No results are available yet that could compare with figure 1 in F, as the spontaneous magnetization has not yet been calculated.

The calculations in F can easily be duplicated for this case. One finds instead of (A. 2)

$$\beta_1 = -1 + 4f,$$
$$\beta_2 = -\tfrac{1}{2} - 6f^{2*},$$
$$\beta_3 = -\tfrac{1}{3} + 4f^2 + \tfrac{40}{3}f^3 - 2f^4,$$

and hence instead of (A. 10)

$$\gamma_1 = 2\xi^2 - \tfrac{2}{3}\xi^3 + \tfrac{1}{6}\xi^4 - \tfrac{1}{30}\xi^5 + \ldots,$$
$$\gamma_2 = \qquad 4\xi^3 - 3\xi^4 + \tfrac{7}{6}\xi^5 + \ldots,$$
$$\gamma_3 = \qquad\qquad 8\xi^4 - 8\xi^5 + \ldots,$$
$$\gamma_4 = \qquad\qquad\qquad 8\xi^5 + \ldots.$$

---

* There is a misprint in F at this point in the sign given.

Substituting in (A. 13) one can obtain a sequence of values for ξ which should converge toward the exact value mentioned above. Listing these values side by side with the corresponding ones in the case of the body-centred cubic lattice we get

square net	body-centred cubic
$\xi_1 = -1$	$\xi_1 = -0\cdot5$
$\xi_2 = -2$	$\xi_2 = -0\cdot5858$
$\xi_3 = -1\cdot3556$	$\xi_3 = -0\cdot5747$
$\xi_4 = -1\cdot3019$	$\xi_4 = -0\cdot6104$
$\xi_5 = -1\cdot3523$	$\xi_5 = -0\cdot5979$

while in reality $\quad \xi = -1\cdot7628 \qquad\qquad \xi = ?$

This shows that signs of convergence in the sequence are not trustworthy and that the ξ obtained by F may be as much as 25 % in error. In the case of the specific heat, the situation is even more serious. Curves of the type plotted in F, figure 2, are very commonly obtained by the use of approximate procedures (as shown in KW figure 8) without being indicative of the actual facts. Whether the singularity is of the type postulated in F is thus still an open question. The only conclusion which is certain is that for other than 50 % composition the specific heat must remain finite because the low temperature curve is independent of composition.

REFERENCES

Born, M. & Fuchs, K. 1938 *Proc. Roy. Soc.* A, **166**, 391.
Fuchs, K. 1942 *Proc. Roy. Soc.* A, **179**, 340.
Kramers, H. A. & Wannier, G. H. 1941 *Phys. Rev.* **60**, 252, 263.

On the statistics of binary systems

By KLAUS FUCHS

University of Birmingham

1. In my paper (Fuchs 1942, in the following quoted as F), criticized in the preceding remarks of Wannier, I was mainly concerned with the statistical properties for varying concentrations. For this reason I omitted to consider some special types of critical points, which are obtained in exceptional cases. I hope I shall be pardoned for omitting to mention these exceptional cases, since they went rather beyond the scope of my paper, and at the time I had no reason to suppose that they were of special interest. However, in order to reconcile the results of Kramers & Wannier

K. Fuchs

(1941, in the following quoted as KW) with my own, it is necessary to assume that one of these special cases represents the correct description of the critical region.

2. In comparing the results of the two theories a clear distinction must be made between facts which have been rigorously proved by the one theory or the other, and those which have only been obtained by approximate evaluation. I shall confine myself first to the former, since this is the only way of settling the points in question.

Rigorous results are given by both theories on the question of a discontinuity in the heat content and the specific heat at the critical temperature, which I shall briefly summarize. For brevity I shall refer to a continuous specific heat, if

$$\Delta c_v = \lim_{\epsilon \to 0} \{c_v(T_c - \epsilon) - c_v(T_c + \epsilon)\}$$

vanishes. This does not necessarily imply that c_v is finite.

3. The results of KW for two dimensions are:

(a) If there is only one critical temperature, then there exist two possibilities; the heat content and the specific heat are either both continuous or both discontinuous.

(b) If there are two critical temperatures, then there exist three possibilities; either the heat contents and the specific heats at both critical points are continuous, or the heat contents are both continuous and the specific heats both discontinuous, or the heat contents are both discontinuous and at least one of the specific heats is discontinuous.

4. The results of my theory are:

(a) The heat content is continuous.

(b) The specific heat is discontinuous, unless either the specific heat is infinite, or the curvature of the heat content (as a function of the concentration) vanishes; in the latter case also the fourth derivative of the free energy vanishes, in addition to the curvature, which vanishes in any case. (If the fourth derivative did not vanish, the two-phase region would end at the critical point in an infinitely narrow peak, a possibility which can probably be excluded.)

If there are two critical temperatures, the same results hold for the higher critical temperature and similar results for the lower one.

5. Combining the results of the two theories one finds:

(i) The heat content is continuous under all circumstances.

(ii) If there is only one critical temperature, the specific heat is also continuous; either it is infinite, or the curvature of the heat content and the fourth derivative of the free energy vanish.

(iii) If there are two critical temperatures, the specific heat is either discontinuous at both temperatures, or continuous at both temperatures; in the latter case either c_v is again infinite, or the curvature of the heat content and the fourth derivative of the free energy vanish.

It should be noted that the theory of KW applies only to two dimensions. In the case of three dimensions the results may be different and only those listed under 4 can be stated with certainty.

6. I wish to indicate now how in my theory the special cases arise in which the specific heat is continuous. It will be convenient to employ the free energy A and the heat content E, rather than the G-functions, or better $a = A/NkT$ and $e = E/NV$ The free energy and the heat content are symmetric in c and $(1-c)$ apart from a linear term which arises from the difference in the energy of the pure substances. The latter is immaterial for this discussion and may be omitted. The equation F (4·3), which determines the limit of solubility $\bar{c}$ may then be written in the form

$$\frac{\partial a}{\partial c} = 0 \quad \text{for} \quad c = \bar{c}. \tag{1}$$

The temperature dependence of $\bar{c}$ is given by F (9·8); i.e.

$$\frac{\partial \bar{c}}{\partial \xi} = -\left[\frac{\partial e/\partial c}{\partial^2 a/\partial c^2}\right]_{c=\bar{c}}. \tag{2}$$

Similarly the specific heat inside the two-phase region is obtained from F (9·12) and (9·13)

$$c_v = c_v(\bar{c}) + Nk\xi^2\left[\frac{(\partial e/\partial c)^2}{\partial^2 a/\partial c^2}\right]_{c=\bar{c}}. \tag{3}$$

In these equations the value of a function for $c = \bar{c}$ should always be taken by approaching the phase boundary in the region of solubility.

The equations can of course be obtained simply by application of the two phase rule. But the derivation given in my paper is independent of the assumptions necessary for the application of that rule. The only assumptions made are that the change in volume with the concentration can be neglected, and that the 'cluster sums' do not depend on the concentration.

At the critical point one has

$$\frac{\partial^2 a}{\partial c^2} = 0, \quad \frac{\partial^3 a}{\partial c^3} = 0, \quad \frac{\partial e}{\partial c} = 0. \tag{4}$$

With (4) it now follows immediately from (3) that c_v is discontinuous at the critical point, unless $c_v(\bar{c})$ is infinite or $\partial^2 e/\partial c^2$ vanishes. In the latter case it follows from (2) that $\partial^4 a/\partial c^4$ vanishes, unless also $\partial \bar{c}/\partial \xi$ vanishes.

7. An exhaustive discussion of these special cases is beyond the scope of these remarks. An infinite specific heat gives rise to an interesting singularity; however, little can be said about it without further study. On the other hand, if the fourth derivative of the free energy vanishes, there arises the possibility of a second critical temperature, similar to that studied by Mayer & Harrison (1938) for the critical region between the gaseous and liquid phase.

According to the general theory the phase boundary is either given by a singularity in the functions $G(c, \beta)$ or by the condition $G_1(\bar{c}, \beta) = 1$, which we may write in the form

$$\left[\frac{\partial^2 a}{\partial c^2}\right]_{c=\bar{c}} = 0. \tag{5}$$

Let us assume that the G-functions (and hence a) is regular below the critical temperature, so that $\bar{c}$ is given by (5). On the other hand, it is also given by (1); it follows that $\bar{c}$ is a double root of the equation (1); at the critical point the two double roots for the two phase boundaries coincide, so that (1) has a fourfold root at the critical point. This implies that the fourth derivative of the free energy vanishes.

Conversely, if the fourth derivative of the free energy vanishes, it is possible that the phase boundary is given by (5). In that case there must exist a second critical temperature, since it can be shown that (5) does not hold for low temperatures.

8. I agree with Wannier's warning against accepting the results of numerical approximation at their face value, as far as values near the critical point are concerned. If a function is evaluated near, or at a singularity by means of an expansion, there exists always the danger that the higher terms will eventually give a contribution; the numerical results of my paper are certainly not sufficient to decide between the alternative possibilities which follow from the exact results.

The position is somewhat different at low temperatures. There we have a splitting into two phases with a large potential barrier. It is likely that in this case a partition function of the homogeneous mixture can be defined, which can be expected to be analytical throughout. It will coincide with the total partition function in the regions of solubility. The 'singularity' in the free energy and the G-functions, is then given by the joint between the two branches of the total partition function; but each branch separately has no singularity.

The discrepancy in the values for the critical temperature may be due to insufficient accuracy of my numerical results. Alternatively, if there are two critical temperatures, the figure of KW must be intermediate between the two. Since my figure corresponds to the higher of the two temperatures (or the lower value of ξ), the discrepancy is in the right direction.

9. The question whether the critical temperature in the case of superstructure should be given by the same equations as in the case of a two-phase region, cannot yet be answered. The equations are derived by approaching the critical point along a phase boundary; this procedure destroys the equivalence of the two cases of equal but opposite interaction energies. The asymmetry in the equations for the critical temperature is therefore no accident.

In the case of a two-phase region, the first equation (4) for the critical temperature follows from (1) and the second from symmetry. These two equations are equivalent with the equations $G_1(c, \beta) = 1$, $G_2(c, \beta) = 0$.

In the case of superstructure the equation (1) does not apply and for the determination of the critical temperature we have to fall back on the general theory.

The latter, however, gives the two equations just quoted only as one alternative; the second alternative is a singularity in the G-functions. If the first alternative is correct, it means that in the case of superstructure the G-functions converge so badly that I failed to find the positive root ξ of the equations (4).

10. In the foregoing discussion I have not included the unpublished results of Onsager to which Wannier refers. Judging from Wannier's remarks they do not appear to contain any statements in contradiction to the above conclusions, though they may conceivably settle some points which have been left open.

REFERENCES

Fuchs, K. 1942 *Proc. Roy. Soc.* A, **179**, 340.
Kramers, H. A. & Wannier, G. H. 1941 *Phys. Rev.* **60**, 252, 263.
Mayer, J. E. & Harrison, S. F. 1938 *J. Chem. Phys.* **6**, 87, 101.

Structure and thermal properties of crystals
V. Thermal expansion of phthalocyanines and porphins

By A. R. Ubbelohde and (Miss) I. Woodward

Measurements have been made on the changes in lattice spacing of various phthalocyanines, over a range of temperatures from 90 to 600° K. Rough visual estimates have also been obtained of changes in the intensities of reflexions with rise in temperature. The molecular movements in the crystals have been calculated from the experimental data. Attention is drawn to the difference in behaviour of hydrogen and platinum phthalocyanines, which is probably due to the increased space required by the platinum atoms as the temperature rises, compared with the hydrogen in hydrogen phthalocyanine.

Apparatus described includes a new design of furnace, suitable for use with various X-ray cameras for single crystal work, and a moving film oscillation camera, which uses the multiple exposure principle to determine small changes in lattice spacing with high relative accuracy.

INTRODUCTION

Measurements of thermal expansion by X-ray powder photographs have in some cases attained a high degree of accuracy. Although there is still considerable scope for extension of the work on simple crystal lattices (*Annual Reports* 1939), powder technique is unsuitable for complex crystals, where the richness of the diffraction pattern frequently makes the interpretation of powder photographs ambiguous. Furthermore, sufficient intensity of reflexion from small spacing planes cannot be obtained from complex crystals in powder form, without very special technique.

In connexion with work on the hydrogen bond in oxalic acid dihydrate (Robertson

& Ubbelohde 1939), attention had been drawn to the interesting information provided by data on the thermal expansions of single crystals of known structure, when this structure is complex. By comparing the thermal expansion in different directions in a crystal, important conclusions can be drawn in certain cases about the molecular movements and forces in the crystal lattice.

The object of the present paper is to describe apparatus, by the use of which all the special photographic methods elaborated for single crystal work can be applied to accurate measurements of the thermal properties of crystals. In order to extend the range of temperatures as much as possible, measurements have been made in the first instance on a number of phthalocyanines and porphins, whose structure had been previously investigated in this laboratory. The conclusions from these measurements are discussed below.

DESCRIPTION OF THE APPARATUS

In order to achieve the flexibility of photographic work which is desirable for measurements on single crystals, it was decided to avoid the construction of instruments in which the high temperature furnace and the precision camera are incorporated in one apparatus. (A number of such instruments are referred to in *Annual Reports* 1939.)

The constant temperature chamber

A small constant temperature chamber was therefore constructed, of total volume about 2·5 c.c., which could be lowered over the crystal after all the necessary setting adjustments had been made, and which could be used with the precision cameras previously described (Robertson 1934, Ubbelohde 1939). The small size of the chamber made it possible to bring the slit system of the collimator within 10 mm. of the crystal, and its low heat capacity enabled the required temperature to be reached within a conveniently short time. Starting from room temperature, the furnace attained a steady state at 400° C within 10–15 min.

Figure 1 gives a diagram of the construction of the chamber and figure 2 shows its general appearance in conjunction with the oscillating moving film camera referred to below. The shell of the chamber consists of a steel cylinder turned down to a thickness of about 0·5 mm. A central segment 5 mm. high is cut away, except for two supporting ribs, to permit free passage of X-rays over approximately 320° of the circle. The top of the cylinder is closed, and is held by a thin metal tube (to minimize heat losses) in a stand with an adjustable arm. The chamber could thus be placed in position over the crystal, and lowered vertically into position by means of a fine screw adjustment. A small plumb line was used to test the vertical alignment.

To improve the uniformity of temperature within this chamber, the segment from which the steel had been cut away was closed by a series of thin plates of metallic beryllium, each 6–7 mm. square, and overlapping slightly on the neighbouring plates.

(These plates were selected from beryllium in the form of 'flitterchen', as supplied by Th. Schuchardt.) A trace of sodium silicate was used to cement the plates to the edges of the steel, above and below the segment cut out. The steel parts of the chamber were next covered with mica, and wound with nichrome wire set in alundum + sodium silicate. Aluminium foil was stuck on the outside of this cement to lessen radiation losses.
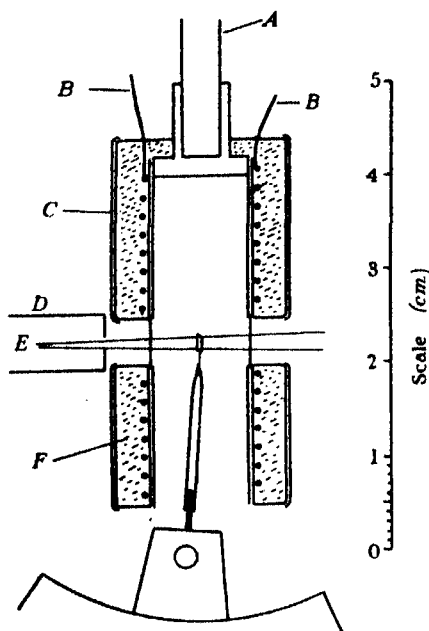
FIGURE 1   FIGURE 2

A. Tubular suspension. *B*. Leads. *C*. Aluminium foil. *D*. Collimator.
E. X-ray beam. *F*. Silicate and alundum cement.
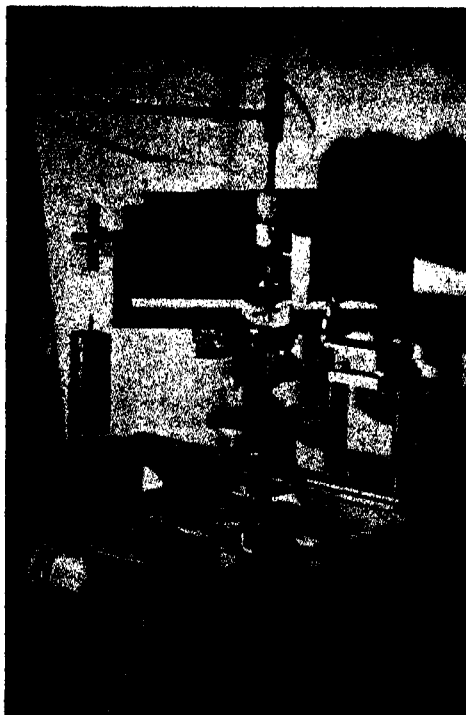
This furnace dissipated about 28 watts at 500° C, and could be used for precise single crystal work even at a distance of about 4 cm. from the X-ray film, without requiring special precautions against film fogging by the radiation. In order to calibrate the chamber, it was heated by the steady current from a set of accumulators, measured on a sub-standard ammeter. The temperature reached with various heating currents was determined, using a calibrated copper constantan thermocouple. The use of a thermocouple was preferred to observations on the melting of known crystals, as it permitted exploration of the temperature gradients within the chamber. The maximum gradient observed near the centre of the furnace, when this was around 400° C, was 0·5° C/mm. On removing the furnace, and then replacing it, the temperature was reproduced to ± 1° C. In fact, this small all metal

chamber behaved like a black body enclosure. Provision was originally made for closing the orifice at the bottom with a loose lid, after inserting the crystal, but this was found to be unnecessary within the limits of temperature indicated.

In measuring the temperature of the air space within the chamber, the hot junction of the thermocouple could be placed at the position subsequently occupied by the crystal; some uncertainty might however arise owing to conduction of heat by the leads. This was investigated by placing various lengths of the coiled leads within the chamber before bringing them outside, and verifying that the temperature recorded by the hot junction was not appreciably affected by the length of the coil. Further tests were made by heating the thermocouple leads outside the chamber with an auxiliary 'guard ring' furnace, and observing that various guard ring temperatures had little effect on the temperature recorded by the hot junction. Thus the true temperature of the air space inside the chamber was being measured. The temperature in this air space was found to be approximately proportional to the wattage dissipated.

Acknowledgement is due to the late Mr C. H. Jenkinson for his help in achieving the compactness of this constant temperature chamber. By using it, thermal expansions of single crystals have been measured over the range from room temperature up to 400° C. Higher temperatures could have been used, but above 400° C the phthalocyanine crystals sublimed too rapidly to permit accurate photography with the X-ray intensities available.

For thermal expansions below room temperature, a design to give corresponding precision has not yet been put into execution, owing to current conditions. The low temperature measurements recorded below were made, using the dropping liquid oxygen technique previously described (Robertson & Ubbelohde 1939). To protect the crystals, these were sealed in cellophane tubes (Lonsdale & Smith 1941). In some cases, these tubes were observed to fill with liquid oxygen; this verified that the temperature of the crystals must have been around $-180°$ C, though the accuracy of this temperature cannot be closely assessed.

An oscillation moving film camera

In addition to the moving film and multiple exposure cameras referred to above, a modification of the moving film camera was constructed, in which the crystal and film could be oscillated in phase through a range of angles from five to fifteen degrees, instead of rotating through the whole 360°. This had the advantage of shortening the exposure, since only selected reflexions in the zero layer line were recorded. Apart from changes in the gearing, no new mechanical feature was involved in the design of this camera. A general view is given in figure 2.

The general principles in applying this camera to measurement of thermal expansions correspond with those previously described for the multiple exposure camera (Ubbelohde 1939). A series of exposures could be taken on the same film, alternately at high temperatures and at room temperatures. Small relative movements of reflexions on this film, due to thermal expansion, could be measured with an

accuracy of 0·05 mm. As previously described, possible errors in the mechanical adjustments of the camera were tested by direct photographic means, and were found not to give any spurious displacements of reflexions, greater than the experimental errors of film measurement. The reflexions were calibrated as previously described, and variations in film shrinkage, as determined with a platinum substandard, were found to be negligible.

With this type of camera, the definition of the reflexion on the film is a function of the Bragg angle. The accuracy of measurement depends on the shape, as well as on the actual displacement of a reflexion due to thermal expansion. This was taken into consideration in selecting a suitable inclination of the film to the primary X-ray beam. In addition to the tables giving the positions of the reflexions in terms of this inclination, it was found convenient to draw up tables giving $\delta d/d$ in terms of $\sin \theta$, for various inclinations of the film holder.

Mounting and setting of the crystals

Some care was required to make mountings sufficiently robust to withstand temperature changes without an accompanying change in setting of the crystals. For work at low temperatures a solution of medium viscosity of cellulose acetate in amyl acetate, or alternatively of bakelite varnish in acetone, could be used to cement the crystal to the supporting fibre. Above room temperature, bakelite solution could be used up to about 300° C; above this temperature, it was necessary to use a silicate dental cement, which was less easy to manipulate. Supporting fibres with negligible absorption for X-rays were made by drawing down one end of thin pyrex tubing to a diameter of about 0·02 mm., and cementing a copper wire into the other end, for fixing on the arcs.

EXPERIMENTAL RESULTS

To investigate the scope of the single crystal measurements, the thermal expansion of a number of phthalocyanines and related crystals was selected for investigation: (1) in view of the large temperature range over which measurements could be made; (2) in view of the comparative information to be obtained from a number of closely related structures, in which the mass of the molecule could be varied considerably by changing the central atom, without affecting the general character of the crystal forces. Measurements were made on hydrogen and copper phthalocyanines, whose structures closely resemble that of the nickel compound (Robertson 1936, Robertson & Woodward 1937), and also on the related crystals of different cell dimensions, platinum phthalocyanine and tetrabenzmonazoporphin (Robertson & Woodward 1940, Woodward 1940).

Intensities of X-ray reflexions

A practical difficulty with these crystals was the 'fading' of reflexions with $\sin \theta > 0·5$, even at temperatures as low as 200° C. In consequence, the largest dis-

placement of reflexions which could be recorded on the films did not exceed 2 mm. Detailed analysis of the temperature factors for various reflexions is of considerable interest, on account of the information such an analysis might give about the amplitudes of atomic vibrations in various directions in a complex crystal structure (*Annual Reports* 1939, p. 156). On account of the additional labour involved, exact intensity measurements have not been attempted in connexion with the experiments on thermal expansion described, in the present paper. Using $I_r = I_0 e^{-2M}$, where $M = 8\pi^2 \sin^2\theta \, \bar{\mu}_x^2/\lambda^2$, the ratio I_{r_2}/I_{r_1} has been calculated from rough visual estimates of the fading, for the temperature ranges indicated below. From ten to twenty observations were made on different reflexions for each of the compounds studied. The random errors obliterate any directional effects which may be established when more accurate intensity measurements are available.

In calculating the following amplitudes, no distinction has been made between the different contributions from different atoms to the reflexions. In the case of the platinum phthalocyanine, the amplitude can be regarded as substantially that of of the platinum atom, in view of its dominant contribution to the reflexions. With the other compounds, the amplitude represents an average for the various atoms in the molecule.

TABLE 1. AMPLITUDE OF ATOMIC VIBRATIONS CALCULATED FROM FADING (I_{r_2}/I_{r_1})

compound	$T_1{}^\circ \text{K}$	$T_2{}^\circ \text{K}$	$\sqrt{(\bar{\mu}_{x_2}{}^2 - \bar{\mu}_{x_1}{}^2)}$ (in A)
HPh	290	475	$0 \cdot 09 \pm 0 \cdot 02$
	290	600	$0 \cdot 12 \pm 0 \cdot 02$
CuPh	290	600	$0 \cdot 08 \pm 0 \cdot 02$
PtPh	290	475	$0 \cdot 11 \pm 0 \cdot 02$
tetrabenzmonazoporphin	290	600	$0 \cdot 13 \pm 0 \cdot 02$

Linear expansions of the crystals on heating

On heating, the crystals of the phthalocyanines became brittle, and thinned by sublimation, particularly normal to the (001) plane. Owing to the fact that the crystals form long thin laths, and owing to the fact that strong reflexions of small spacing occur only for low values of k, measurements were most accurate in the case of oscillations about the b axis. Measurements in directions lying outside the ac plane were made only in the case of hydrogen and platinum phthalocyanine.

The crystals of tetrabenzmonazoporphin did not appear to become brittle on heating.

The large number of measurements made on the linear expansions of the crystals can be compactly presented in the form of table 2, giving the increments of the principal axes of the expansion ellipsoid.

Data were insufficient for computation of the probable errors in α_{22} according to standard theory. A rough estimate is $10^4 \alpha_{22} = 1 \cdot 14 \pm 0 \cdot 1$ for Pt phthalocyanine.

The following comments may be made on the results given in table 1 and table 2.

(1) *Relation between the expansions of various crystal planes.* In a monoclinic

crystal, one of the principal axes of the expansion ellipsoid coincides with the b axis, and the positions of the remaining axes α_{11}, α_{33} in the (010) plane have to be determined experimentally. If ψ is the angle between the crystallographic c axis and the direction of the principal expansion axis α_{11}, the expansion α' in any direction making an angle ξ with the c axis is given by

$$\alpha' = A + B \cos 2\xi + C \sin 2\xi,$$

where $\quad \tan 2\psi = C/B; \quad \alpha_{11} = A + B/\cos 2\psi; \quad \alpha_{33} = A - B/\cos 2\psi.$

The values of A, B and C in the above were calculated from experimental values of α'. Each separate reflexion was measured on one or more independent films; the number of independent observations used to establish the line of closest fit is given in the first column of table 2. Standard probability theory gave the errors in A, B and C. The probable errors in α_{11}, α_{33}, and ψ cannot be related to the probable errors in A, B and C by simple theory. Limiting values greater than the actual probable errors are therefore inserted in table 2.

TABLE 2. MEAN EXPANSION/°C ALONG THE PRINCIPAL
AXES OF THE EXPANSION ELLIPSOID

substance	series	number of observations	temp. range (°K)	$\alpha_{11} \times 10^4$	$\alpha_{33} \times 10^4$	ψ	$\alpha_{22} \times 10^4$
hydrogen phthalocyanine	1	51	90–290	0·65 ± 0·03	0·24 ± 0·03	115·0° ± 3·0	—
	2	33	290–475	0·78 ± 0·04	0·40 ± 0·04	109·1° ± 7·1	+ 0·18
	3	21	290–600	0·75 ± 0·02	0·40 ± 0·02	109·2° ± 2·2	—
copper phthalocyanine	4	19	90–290	0·65 ± 0·07	0·28 ± 0·07	95·4° ± 6·5	—
	5	18	290–600	0·73 ± 0·04	0·31 ± 0·04	101·1° ± 4·7	—
platinum phthalocyanine	6	23	290–475	0·60 ± 0·05	− 0·14 ± 0·05	73·8° ± 3·5	1·14
tetrabenzmonazoporphin	7	27	290–670	0·93 ± 0·03	0·14 ± 0·03	99·4° ± 1·2	—

(2) *Effect of temperature on thermal expansion.* Comparison of the expansion for any one crystal over various temperature ranges shows that some falling off in α_{11} and α_{33} takes place between 290 and 90° K. If the Debye Waller theory is applied to the fading in intensities with rise of temperature, the θ temperature of all the crystals can be calculated to be of the order of 180° K. With such complex crystals, the theory is probably only a rough approximation. It does, however, indicate that some falling off in expansion coefficients should occur below 180° K, in agreement with the experimental observations.

(3) *Correlation of overall expansion of the crystals with changes in the space occupied by the molecules.* It was shown by Robertson (1936, p. 1203) that the orientation of the phthalocyanine molecule, the position of whose centre is determined by the crystal symmetry, can be conveniently described by the position of three mutually perpendicular lines L, M, N in the molecule, the 'molecular axes'. L and M lie in the plane of the molecule, and the coordinates of the atoms referred to these axes show approximately a fourfold symmetry.

A general picture of the relationship between the individual molecules and the crystal expansion can be obtained from inspection of the scale drawings, figures 3, 4, 5 and 6, which give the expansions in the directions shown, as calculated from the known maxima α_{11}, α_{22}, α_{33}, and the known value of ψ. There is a marked change in the behaviour on heating of the phthalocyanine molecules, when the central hydrogen atoms are replaced by platinum. (Compare figure 3 with figure 4, and figure 5 with figure 6.)

Detailed analysis of the difference would have to be based on a complete correlation of the changes in cell dimensions on heating, with the expansions δL, δM, δN per unit length of the molecular axes, and changes in the orientation of the molecules by rotations $\delta\theta_L$, $\delta\theta_M$, $\delta\theta_N$ about L, M and N. Since one of the parameters of the expansion ellipsoid in the case of a monoclinic crystal is determined by crystal symmetry, only five equations are available to determine these six quantities. From the details of a typical calculation (Appendix 1) it will be seen that although no unique solution is possible, some of these quantities can be determined directly, and others by making reasonable assumptions. Those results which can only be calculated after making the assumptions described in Appendix 1 are placed in brackets in table 3.

TABLE 3. MOLECULAR MOVEMENTS/°K ($\times 10^{+4}$)

	hydrogen phthalocyanine	platinum phthalocyanine
δL	0·51	0·55
δM	(0·51)	(0·55)
δN	(0·34)	(0·52)
$\delta\theta_L$	(0·15)	(−1·34)
$\delta\theta_M$	0·33	0·75
$\delta\theta_N$	−0·10	−0·15

Attention may be drawn to two points of interest for the relation between structure and thermal properties:

(1) The increase in space required by the molecules at higher temperatures, as calculated from the linear expansions in table 3, agrees with the increase in amplitude of atomic vibrations, as calculated from intensities (table 1). For example, in the case of PtPh, $\delta L = 0.55 \times 10^{-4}/°K$, or over the temperature range 290 to 475°K $\delta L = 1.02 \times 10^{-2}$. L is approximately 12·5 A, so that the increased 'size' of the molecule in this direction is $1.02 \times 10^{-2} \times 12.5\,A = 0.13\,A$. This may be compared
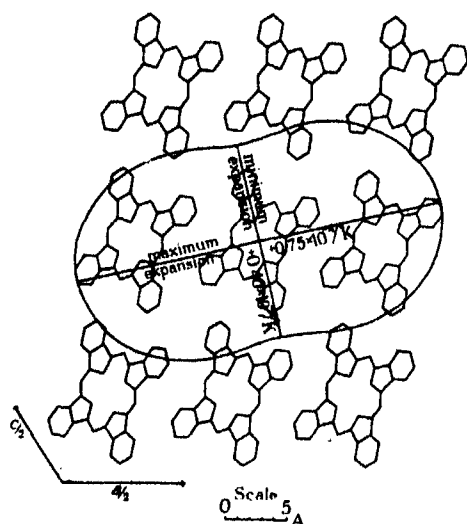
FIGURE 3. Hydrogen phthalocyanine. Projection on (010) plane, showing average thermal expansion per °K from 290 to 600° K.
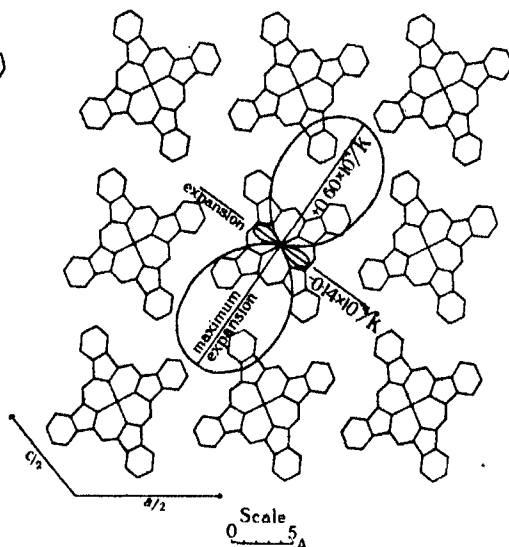
FIGURE 4. Platinum phthalocyanine. Projection on (010) plane, showing average thermal expansion per °K from 290 to 475° K.
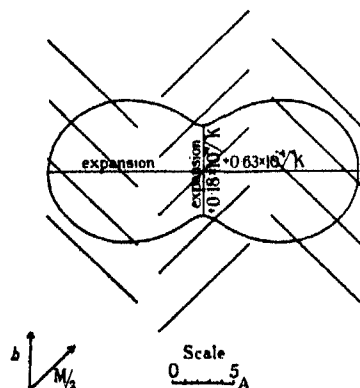
FIGURE 5. Hydrogen phthalocyanine. Average expansion per °K in plane containing *b* crystal axis and *M* molecular axis, showing effect on slope and separation of the molecules.
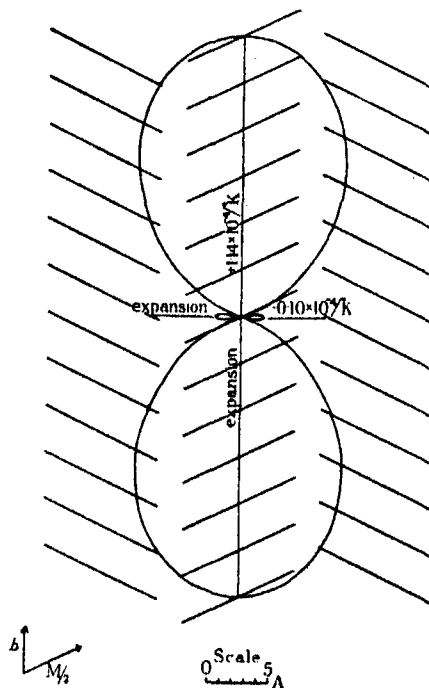
FIGURE 6. Platinum phthalocyanine. Average expansion per °K in plane containing *b* crystal axis and *M* molecular axis, showing effect on slope and separation of the molecules.

with an average amplitude of 0·11 A (table 1) derived from intensity measurements.

(2) The differences between platinum and hydrogen phthalocyanine can be ascribed to the space requirements of the platinum atoms.

Replacement of the hydrogen by copper makes some difference in α_{33}, but this, is only just outside experimental error (cf. table 2). Restricting the comparison to hydrogen and platinum phthalocyanines, the main difference in the structures arises from the fact that the molecules, which are practically identical in size and shape, are differently inclined to the (010) plane. For PtPh, the angle of inclination is 26·5°, as compared with 44·2° for HPh. This difference in inclination is probably due to the larger spatial requirements of the Pt atom perpendicular to the plane of the molecule.

The differences in the molecular movements on heating correspond with this difference in structure. Whereas the values of δL, δM and δN are of the same order, the changes in orientation of the molecules are such as to diminish the inclination to the (010) plane for HPh, and increase it for PtPh. The effect is to increase the distance of closest approach between nitrogen and carbon atoms in parallel molecules, in HPh (approx. $1·4 \times 10^{-4}$ A/°K). In PtPh there is a greater increase in the distance of closest approach between C and N atoms in parallel molecules, and also in the distance between the platinum atoms (approx. $4·3 \times 10^{-4}$/°K) and between platinum and N atoms (approx. $2·8 \times 10^{-4}$/°K). This increased space requirement imposed by the platinum atoms appears to be an important thermal effect in the crystal.

Acknowledgements are due to the Managers of the Royal Institution for the facilities given for carrying out this investigation, and to Professor R. P. Linstead for supplying specimens of the crystals used. We should like also to express our grateful appreciation of the interest shown by the late Sir William Bragg in the progress of the work.

APPENDIX 1

Calculation of the molecular movements from the expansion of the crystals. The method adopted is to equate changes in the magnitude and directions of the projections of the molecular axes, L, M, with the crystal expansions projected on the ac plane. The centre of the standard molecule may be taken as origin, 0. As axes in the ac plane it is convenient to take the projection of L, and a line perpendicular to it, which lies 2·2° from the projection of M (figure 7).

Points at unit distance from the origin along the L and M axes are considered. Their projections before thermal expansion may be denoted by P and Q, and after expansion by P' and Q'. The coordinates of these points in the ac plane may be written

$$x_P, z_P; \quad x_Q, z_Q; \quad x_{P'}, z_{P'}; \quad x_{Q'}, z_{Q'}.$$

Taking the expansion of HPh from 90 to 290° K for a specimen calculation,

$$10^4 \alpha_{11} = 0.65/°\text{K}; \quad 10^4 \alpha_{33} = 0.24/°\text{K}; \quad \psi = 115° = \beta - 7.2°.$$

From the known structure of HPh (Robertson 1936, p. 1203) the projections of OP on the a axis and on the perpendicular to it in the ac plane are 0.3532 and 0.9347. It follows that $OP = 0.9992$ at 69.3° to the a axis, i.e. at 62.1° to α_{11} (figure 7).
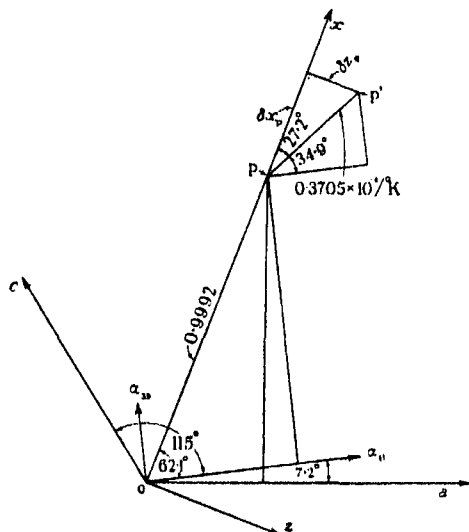
FIGURE 7. Diagram illustrating Appendix 1.

Hence   projection of OP on $\alpha_{11} = 0.9992 \cos 62.1° = 0.4675$,

and   projection of OP on $\alpha_{33} = 0.9992 \sin 62.1° = 0.8831$.

Multiplying by α_{11} and α_{33} respectively,

$$10^4 \times \text{projection of } PP' \text{ on } \alpha_{11} = 0.65 \times 0.4675 = 0.3039/°\text{K},$$
$$10^4 \times \text{projection of } PP' \text{ on } \alpha_{33} = 0.24 \times 0.8831 = 0.2119/°\text{K}.$$

Thus $PP' = 0.3705$ at an angle of 34.9° with α_{11}, i.e. 27.2° with OP. Therefore

$$\delta x_P = x_{P'} - x_P = 0.33 \times 10^{-4}/°\text{K},$$
$$\delta z_P = z_{P'} - z_P = 0.17 \times 10^{-4}/°\text{K}.$$

A similar calculation for Q shows that

$$\delta x_Q = 0.11 \times 10^{-4}/°\text{K},$$
$$\delta z_Q = 0.44 \times 10^{-4}/°\text{K}.$$

Thus from the observed crystal expansions, the following table may be drawn up:

temperature range	$10^4 \delta x_P /°\text{K}$	$10^4 \delta z_P /°\text{K}$	$10^4 \delta x_Q /°\text{K}$	$10^4 \delta z_Q /°\text{K}$
90/290° K	0.33	0.17	0.11	0.40
290/600° K	0.51	0.16	0.10	0.46

The values of δx_P, etc. can also be calculated in terms of the molecular movements:

The standard and reflected molecules in the HPh structure must be taken together, as in some cases their movements give an additive shift, and in other cases they cancel. The separate contributions may be listed as follows:

(a) δL (effects additive). L is inclined to $2 \cdot 3°$ to Ox and is perpendicular to Oz. Hence the contributions are

$$\delta x_P = 1 \cdot 00 \delta L, \quad \delta z_P = \delta x_Q = \delta z_Q = 0.$$

(b) δM (effects additive). M is inclined at $44 \cdot 1°$ to the ac plane and its projection makes an angle of $-2 \cdot 2°$ with the z axis. Hence the contributions are

$$\delta x_P = \delta z_P = 0, \quad \delta x_Q = -0 \cdot 03 \delta M, \quad \delta z_Q = 0 \cdot 72 \delta M.$$

(c) δN (effects cancel).

(d) $\delta \theta_L$ (effects additive). For this calculation a small change ϵ was made in $\cos \psi_M$ (Robertson 1936, p. 1203), $\cos \chi_M$ and $\cos \omega_M$ being changed by $k_1 \epsilon$ and $k_2 \epsilon$, so that M remained perpendicular to L. Neglecting ϵ^2, it was found that Q underwent a displacement making an angle of $-18 \cdot 4°$ with the a axis. After correlating ϵ with $\delta \psi_M$ and so with $\delta \theta_L$, the contributions are found to be

$$\delta x_P = \delta z_P = 0, \quad \delta x_Q = 0 \cdot 03 \delta \theta_L, \quad \delta z_Q = 0 \cdot 70 \delta \theta_L.$$

(e) $\delta \theta_M$ (effects additive). Using a similar procedure to (d), the contributions are found to be

$$\delta x_P = -0 \cdot 03 \delta \theta_M, \quad \delta z_P = 0 \cdot 69 \delta \theta_M, \quad \delta x_Q = \delta z_Q = 0.$$

(f) $\delta \theta_N$ (effects additive). Here PP' is parallel to the projection of M and equal to $0 \cdot 72 \delta \theta_N$ and QQ' is parallel to the projection of L and equal to $-1 \cdot 00 \delta \theta_N$. Hence the contributions are

$$\delta x_P = -0 \cdot 03 \delta \theta_N, \quad \delta z_P = 0 \cdot 72 \delta \theta_N, \quad \delta x_Q = -1 \cdot 00 \delta \theta_N, \quad \delta z_Q = 0.$$

Summing these independent displacements to obtain the resultant values,

$$\delta x_P = 0 \cdot 33 \times 10^{-4} = \quad 1 \cdot 00 \delta L - 0 \cdot 03 \delta \theta_M - 0 \cdot 03 \delta \theta_N,$$
$$\delta z_P = 0 \cdot 17 \times 10^{-4} = \quad\quad\quad\quad 0 \cdot 69 \delta \theta_M + 0 \cdot 72 \delta \theta_N,$$
$$\delta x_Q = 0 \cdot 11 \times 10^{-4} = -0 \cdot 03 \delta M + 0 \cdot 03 \delta \theta_L - 1 \cdot 00 \delta \theta_N,$$
$$\delta z_Q = 0 \cdot 40 \times 10^{-4} = \quad 0 \cdot 72 \delta M + 0 \cdot 70 \delta \theta_L.$$

In addition to these four equations, giving the molecular movements in terms of the expansion of the crystal, a fifth equation is obtained from the expansion in the direction of the b axis. Owing to the relation between the standard and reflected molecules, δL, δM and $\delta \theta_N$ make no contribution to this expansion, and for the mean expansion/° K

$$\alpha_{22} = \delta N - 0 \cdot 97 \delta \theta_L - 0 \cdot 03 \delta \theta_N.$$

From the first four equations, solutions are obtained for δL, $\delta \theta_M$ and $\delta \theta_N$ for the temperature range 90 to 290° K, similar to those given for the range 290 to 600° K

in table 3 above. In addition, a numerical value is obtained for $\delta M + \delta\theta_L$. Now the arrangement of the molecules in PtPh is such as to suggest $\delta L = \delta M$. Adopting this hypothesis, it is possible to solve for $\delta\theta_L$, and hence it is also possible to obtain a value for δN from the fifth equation. Values dependent on this hypothesis are given in brackets in table 3. Values obtained for HPh on the same hypothesis are also given in table 3. Actually the structure of HPh is somewhat different from that of PtPh, and an alternative hypothesis would be to put δM intermediate between δN and δL, e.g. $\delta M = \frac{1}{2}(\delta N + \delta L)$. This does not make much difference to the figures in brackets in table 3, and thus verifies that these are likely to be near the truth.

REFERENCES

Ann. Rep. Chem. Soc. 1939.

Lonsdale, K. & Smith, H. 1941 *J. Sci. Instrum.* **18**, 133.

Robertson, J. M. 1934 *Phil. Mag.* **18**, 729.

Robertson, J. M. 1936 *J. Chem. Soc.* p. 1195.

Robertson, J. M. & Ubbelohde, A. R. 1939 *Proc. Roy. Soc.* A, **170, 222.**

Robertson, J. M. & Woodward, I. 1937 *J. Chem. Soc.* p. 219.

Robertson, J. M. & Woodward, I. 1940 *J. Chem. Soc.* p. 36.

Ubbelohde, A. R. 1939 *J. Sci. Instrum.* **16**, 155.

Woodward, I. 1940 *J. Chem. Soc.* p. 601.

INDEX TO VOLUME 181 (A)

ERRATUM

On page 210, line 5, *for* Bermann *read* Bergmann.